INVESTIGATION OF THE EFFECT OF LOW MOLECULAR WEIGHT PEG ON LYSOZYME INTERACTIONS IN SOLUTION USING COMPOSITION GRADIENT STATIC LIGHT SCATTERING

THESIS

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Science in the Graduate

School of The Ohio State University

By

Shikha Kantilal Gandhi, B.E.

The Ohio State University
2008

Examination Committee:

Professor Michael Paulaitis, Advisor

Professor Barbara Wyslouzil

Approved by:

_____

Advisor
Graduate Program in Chemical Engineering

ABSTRACT

Conventional approaches towards protein crystallization are based on predicting favorable solution conditions for crystal formation through trial and error and are expensive. The second virial coefficient (protein-protein), a thermodynamic solution parameter, is a measure of protein interactions in solution, and more importantly-experimentally accessible. It allows a systematic approach for predicting solution conditions favorable to protein crystallization based on solution thermodynamics.

Poly-ethylene glycol (PEG) is a commonly used salting-out agent in protein crystallization. Its effect on influencing protein-protein interactions in solution is attributed popularly to a “depletion effect” in the literature. This work provides experimental evidence that low molecular weight PEG induces repulsion between protein molecules, contrary to the predictions of depletion, if a third protein-protein-polymer virial coefficient is considered in the analysis, in addition to protein-protein and protein-polymer second virial coefficients.

DEDICATION

Dedicated to my father

ACKNOWLEDGEMENTS

I thank my advisor, Dr. Michael Paulaitis, for his guidance and support throughout my program. I've benefited tremendously from his disciplined approach towards problem-solving, both in my research, and as an individual. I also thank him for his consideration of my special circumstances with my husband being away and giving me the opportunity and flexibility to balance things at home.

I thank all my research group members, for always keeping things fun at work. A special note of thanks to Hamsa, for all the patient "explanatory sessions", it would have taken me a lot longer to grasp all those concepts on my own!

I thank my family for all the moral support, love, and comfort that they have always provided me. They have always inspired me to give my best to everything I do and I love them very much!

Last, but not the least, I thank my husband, Tarun, for his patience, perseverance, and sacrifices, without which, I would not have come this far. His unconditional support made all the challenges that come with striking a work-life balance seem easy, although they were far from it. He went out of his way every-time, from waking up at three in the morning to catch a flight every week just so we could be together, to simple but meaningful things, like taking care of the cooking and cleaning when I had too much

work to finish. I appreciate and cherish everything he has done to help me realize my goal.

I have been very fortunate to have met some wonderful people in these two years, and I thank them all, as they have all contributed in making my experience at Ohio State a memorable one!

VITA

December 30, 1983	**Born- Karav, India**
2005	**B.E., Chemical Engineering, Dharmsinh Desai University**
2005-present	**Graduate Teaching and Research Associate, The Ohio State University**

FIELDS OF STUDY

Major Field: Chemical Engineering

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Protein Crystallization and the Second Virial Coefficient

Downstream processing accounts for more than 70% of the total manufacturing costs of therapeutic proteins as the process economics are closely governed by successful design and scale up of separation operations[1]. However, due to the complexity and delicacy of biological systems, most separation processes for protein purification are poorly characterized and hence expensive to scale up[2].

Crystallization is usually the final unit operation in protein purification as most therapeutic proteins are sold in the crystalline form. Crystalline proteins not only offer a higher shelf life compared to in-solution formulations[3], they also offer the ease of convenient oral delivery and can be readily reconstituted to solution for intravenous delivery.

In essence, the key to obtaining any crystal lies in determining the "right" set of solution conditions, for which the solute (proteins in this context) exceeds the solubility limit and comes out of solution as an ordered structure instead of an aggregated precipitate[4]. However, due to lack of proper characterization, these conditions are determined by screening thousands of solutions through trial and error.

Thus, in-spite of being a bottle-neck in manufacturing, the approach to understanding protein crystallization remains more or less empirical and unstructured.

In addition to the biopharmaceutical industry, understanding protein crystallization is extremely important to structural biology as the field relies on successfully crystallizing biological macromolecules to determine their crystal structures through x-ray crystallography, along with structural information obtained through techniques like nuclear magnetic resonance (NMR).

George and Wilson[5] first attempted to develop a systematic approach to protein crystallization in 1994 where they showed a direct correlation between the osmotic 2nd virial coefficient (B_{ii}), a thermodynamic parameter describing intermolecular interactions between protein molecules in solution, and solution conditions favorable to crystal formation. They found that the solution conditions resulting in protein crystallization correspond to B_{ii} values that fall within a range of moderately negative values, the "crystallization slot". Positive values of B_{ii} correspond to repulsive pair-wise interactions between protein molecules and cause no phase separation, whereas large negative values correspond to strong attractive interactions that result in aggregate formation.

By definition, B_{ii} is the 2nd order coefficient in the expansion of osmotic pressure as a function of solute concentration:

$$\Pi = \left(\frac{RT}{M_i} \right) C_i + B_{ii} C_i^{\,2} \quad , \qquad (1.1)$$

where, Π is the equilibrium osmotic pressure of the solution, M_i is the number average solute molecular weight of species i, C_i is the solute concentration on a weight/volume basis, R is the universal gas constant, and T is the absolute temperature.

Equation 1.1 allows for straightforward calculations of B_{ii} by measuring the osmotic pressure as a function of solute concentration. A more detailed derivation of B_{ii} from osmotic pressure is given by Cantor and Schimmel[6] and will be briefly discussed in 2.1.1. Thermodynamically, B_{ii} is related to the potential of mean force, w, between two protein molecules in solution[7,8]. The potential of mean force is defined as the work required to bring two protein molecules (labeled i and j for the sake of clarity) of given shapes and orientations that are infinite distance apart to a finite separation distance, r.

$$B_{ij} = \frac{1}{2}\int_{\Omega_i}\int_{\Omega_j}\left[\frac{1}{3}r_c^{\,3} - \int_{rc}^{\infty}\left(e^{-w/kT}\right)r_{ij}\,dr_{ij}\right]d\Omega_i\,d\Omega_j \ , \qquad (1.2)$$

where, r_{ij} is the distance between the two molecules, Ω_i and Ω_j are the normalized angular vectors, and r_c is the separation distance upon contact between the molecules.

1.2 Factors Affecting B_{ii} Values

B_{ii} is a measure of weak protein-protein interactions in solution on account of electrostatic interactions between the highly charged proteins, hydrophobic interactions due to attraction between hydrophobic sites, and dispersion interactions[8]. In addition to temperature[9] and protein concentration[10] in solution, there are several factors such as pH[10,11], ionic strength[12], and the presence of additives[13] that affect the value of B_{ii}. A brief discussion of these factors and the specific interactions they affect is explained in this section.

1.2.1 Effect of pH

Proteins are a class of (poly)electrolytes[14], i.e., charged macromolecules that dissociate spontaneously in solution to create charged (poly)ions in an atmosphere of counter-ions

along with ions from added salt. When the solution pH decreases below neutral, H^+ atoms bind to the negatively charged COO^- residues in Asp and Glu, neutralizing their –ve charge. In addition, H^+ atoms bind to the vacant electron pairs on the nitrogen of the amino group resulting in a positively charged $–NH_3^+$ residues. Thus, a decrease in pH results in the molecule gaining a net positive charge in addition to changing the local charge distribution on the molecule. Similarly, when the solution pH increases, protons are removed from –COOH groups on Asp and Glu and from $–NH_3^+$ groups in Lys and Arg resulting in $–COO^-$ and $–NH_2$ residues respectively. Hence, with increasing pH, the net charge on the molecule becomes more negative in addition to changing the local charge distribution on the macromolecule. Thus, changes in pH directly influence the nature of electrostatic interactions between protein molecules in solution and hence the B_{ij} values.

1.2.2 Effect of Ionic Strength

The ionic strength, I, of a solution is a measure of the amount of charge present in solution on account of added salt, and is defined as:

$$I = \frac{1}{2}\sum c_i z_i^{\,2} \quad , \tag{1.3}$$

where, c_i and z_i are the concentration and charge of a given ionic species i , respectively. The addition of salt to a protein solution can both increase (salting in) and decrease (salting out) protein solubility depending on the nature of the protein (charge) as well as the concentration and nature of the salt[15]. It has been observed that low amounts of added salt generally increase the protein solubility in solution, whereas high salt concentrations generally have a salting out effect[16]. There are several empirically determined

expressions correlating protein solubility in an aqueous solution as a function of salt concentration, for example, one of the earliest and most widely accepted, Cohn's equation[17] states that the solubility of the protein falls logarithmically with an increase in salt concentration according to:

$$\ln S_{protein} = \alpha - \beta c_{salt}, \tag{1.4}$$

where, $S_{protein}$ is the solubility of the protein , c_{salt} is the molarity of the added salt, α, and β are empirical constants.

Ammonium sulfate, sodium acetate, and sodium chloride are three of the most popular additives (salts) used by protein crystallographers[18,19]. Although sodium chloride and sodium acetate are poor salting out agents on their own[19], they are generally one of the several dominant components of a "salt-cocktail"[20] used to crystallize proteins, and are also the salts of choice (particularly sodium chloride) in theoretical investigations of protein interactions in solution, as their monovalent nature is relatively easier to model requiring comparatively lesser computational time than multi-valent salts[21]. The Hofmeister series[22], proposed in 1888, gives the effectiveness of commonly used anions and cations in "salting out" or precipitating/crystallizing proteins. Generally, protein crystallization requires lower salt concentrations compared to protein precipitation[12].

At low salt concentrations, (< 0.25 M - 0.5 M), electrostatic interactions play an important role in contributing to the PMF, however, at higher salt concentrations, they are completely screened by the added salt ions and other contributions dominate the interaction between protein molecules[12]. In the low-salt regime, B_{ii} values can both

increase or decrease with increasing salt concentration, i.e., attractions between protein molecules can be both, repulsive or attractive, depending on the type (charge distribution) of protein. For example, in the low-salt region, for proteins like myoglobin and cytochrmome-c, B_{ii} values increase with increasing salt concentration, whereas, for proteins like lysozyme, chymotrypsinogen, bovine serum albumin, catalase, and a-lactalbumin, B_{ii} values decrease with increasing salt concentration[12,23,24]. The trend is consistent for both ammonium sulfate as well as sodium chloride. In the high-salt region, B_{ii} is almost constant with increasing salt concentration for sodium chloride and ammonium sulfate, however, a threshold concentration (varying with type of protein) is observed for ammonium sulfate, above which, B_{ii} values begin to decrease systematically[12]. As reviewed by Ruckenstein et al.[16], several correlations between B_{ii} values and protein solubility, as well as their molecular origins, have been proposed and discussed in the literature[25,26,27] for well-known protein-salt systems.

1.2.3 Effect of Additives

In addition to salts, several other additives such as monohydric alcohols[28] (methanol, ethanol, 1-propanol, n-butanol), polyols[13,27] (glycerol), and polymers[29] (polyethylene-glycol), are known to affect protein-protein interactions in solution. However, the mechanisms through which these additives affect protein interactions are not well characterized. This section is a brief literature review of the observed effects and proposed mechanisms of how these additives influence protein interactions in solution, with special reference to PEG.

Prausnitz et al[27]. showed that for lysozyme in constant solution conditions of .05M NaCl and pH 7, B_{ii} approximately doubled in the presence of methanol and ethanol, whereas, it increased by a factor of about 2.5 in the presence of glycerol, compared to B_{ii} values obtained in the absence of any additive. They attribute this increase in B_{ii} value, or the weakening of protein-protein interaction, to a possible adsorption of monohydric alcohol molecules to hydrophobic sites on the protein surface, and thus a decrease in extent of hydrophobic attractions between two protein molecules. However, the adsorption mechanism does not account for the similar and more pronounced effect observed in the presence of glycerol, which owing to its three polar -OH groups is more hydrophilic than its monohydric counterparts, and as a result does not bind to hydrophobic sites on the protein. Instead, as glycerol prefers the polar solvent rather than the protein's hydrophobic sites, it partitions into the solvent resulting in an increased hydration layer around the protein, and affects protein interactions on account of an increased excluded volume of the protein. However, as pointed out by Paliwal et al.[30], the treatment of the hydration layer as one of uniform thickness (roughly 3 Å) and attributing it to an increased excluded volume based on an idealized hard-sphere geometry of the protein is inappropriate, as the distribution of hydration sites is governed by the interaction of water molecules with its local environment, in addition to local effects such as differences in surface roughness and angular geometries of the protein.

Polyethylene glycol is a commonly used additive to induce protein crystallization[31]. Most model systems in the literature consider the polymer and protein to be hard-spheres and attribute the PEG induced attraction between proteins to a "depletion-effect"[31,32,33,34]. As hard-spheres, protein and polymer molecules cannot interpenetrate, and the center of a

polymer molecule is excluded from the region around the protein by a distance the order of the effective radius of the polymer, known as the depletion zone. As the polymer disperses homogenously in the solvent (water), an "overlap" of depletion zones around protein molecules is favored, resulting in an imbalance in osmotic pressure that "pushes" the particles together. This osmotic pressure imbalance is modeled as an attractive pair-wise potential, known as the depletion potential. The depletion effect considers PEG as an "inert" polymer that affects protein solubilities in water on account of its size alone. However, the properties of PEG and its interactions with proteins and solvent (water) depend on its molecular weight as well as concentration[38]. Several groups have studied the exclusion of PEG from the protein surface as a function of PEG molecular weight and concentration. Timasheff et al.[31] report an increase in the preferential hydration, similar to the "uniform" hydration concept used by Prausnitz et al.[28], of β-lactoglobulin with an increase in PEG molecular weight (200, 400, 600, and 1000 Da). They attribute the increase in protein hydration to an increase in the excluded volume of PEG with increasing molecular weight, and hence an "enhanced" exclusion. However, they also observe a decrease in PEG exclusion with an increase in PEG concentration for all molecular weights studied except PEG 200. This "concentration effect" is more pronounced for PEG 1000 than for lower molecular weight PEGs. Based on the fact that PEG is essentially a non-polar molecule, they propose that while the primary mechanism contributing to the decrease of protein solubility by addition of high molecular weight PEG is due to its excluded volume, PEG binds to the protein at high concentrations (of PEG) through hydrophobic interactions. Although, the authors acknowledge that the range of PEG molecular weights studied by them is limited and the exclusion effects for

PEG molecular weights (sizes) that are comparable to that of the protein (>8000 Da) are not considered. The limit of the polymer molecule being smaller than the protein is referred to as the "protein limit" by Kulkarni et al., who report B_{ii} values of proteins in the presence of PEG in a molecular weight range beyond the protein limit as established by them. They conclude their results to be in agreement with the depletion model, with the excluded volume of PEG being the sole factor determining its properties as a protein precipitant. However, their analysis ignores any interactions between the protein and polymer on the assumption that the PEG monomer is much smaller than the protein, which is inappropriate as PEG molecules do not interact with proteins as monomers in solution, but rather large polymers of comparable size to the protein.

Although the depletion effect is widely attributed due to the size of the polymer molecules, there have been alternative mechanisms due to the charges on the protein.

For example, Lee et al.[36] propose that PEG's exclusion from the protein surface is based on unfavorable interactions between PEG and charges on the protein. Ingham et al.[32] report a significant increase in the pH of protein solutions (maximum of .2 units) with the addition of PEG at high concentrations, an observation that could support a change in the charge distribution on the protein in the presence of PEG.

Although different in the basis for depletion, both the size and charge theories contributing to depletion predict the same qualitative behavior, i.e. PEG is an inert molecule and does not directly interact with the protein, however, its addition to a protein solution induces an attraction between protein molecules.

However, several statistical thermodynamic models[36,37,38] have been proposed suggesting an existence of a weak attraction between protein and polymer molecules in addition to

the steric repulsion on account of excluded volume effects. Abbott et al.[39] showed experimentally through neutron scattering measurements that although the net interaction between BSA and PEG was repulsive, the magnitude of the repulsive interaction (in terms of Bij) was smaller compared to one calculated statistically based on purely excluded volume effects, providing experimental evidence that a weak attraction between PEG and BSA was offsetting the theoretical value based on steric repulsion alone. The notion that PEG is an inert impenetrable polymer was further challenged by Sheth and Leckband[40] who showed the presence of attractive interactions between PEG-protein (streptavidin) through compressive force measurements. They explain the attraction through possible changes in the polymer configuration along with a protein penetration of the polymer core, thus directly suggesting that PEG is capable of binding protein.

They attribute their results to PEG's unique properties in water[41], and to the ethylene-oxide chain binding solvent as well as other polymer (protein) molecules. It has been shown that low molecular weight PEGs (<10,000 Da) cause cells and vesicles to adhere to one another (aggregate) in solution, while high molecular weight PEGs cause the cells to remain dispersed in solution[41]. Fraden et al[42] found that low molecular weight PEG (1000 Da) at low concentrations (<100mg/ml) depressed the cloud point of lysosyme solutions, indicating a stabilizing effect in agreement with the depletion model, however, high molecular weight PEG (8000 Da) in the same dilute concentration range increased the cloud point of lysozyme solutions, indicating a destabilizing effect contrary to the depletion effect. At higher PEG concentrations, cloud points always increased independent of the molecular weight of the polymer. Similar results were obtained by

Benedeck et al.[43] with cloud point measurements of γ-crystallin solutions in the presence of PEG.

Based on the experimental evidence available in the literature, it is clear that the mechanism through which PEG influences protein-protein interactions in solution needs further characterization and it is difficult to conclude whether or not PEG-protein interactions contribute to the cannot be modeled according to a depletion effect alone. The investigation of the effect of PEG molecular weight and concentration on protein-protein and protein-polymer interactions in solution is one of the main objectives of this work. The mathematical framework selected to carry out this investigation is as given by Fraden et al.[42] and will be described in subsequent chapters.

1.3 References

1. Aldridge, S. 2006. Downstream processing needs a boost. *Bioprocessing Channels: Feature Articles.* 26:1.

2. Sommerfeld, S., and Strube, J. 2005. Challenges in biotechnology production – generic processes and process optimization for monoclonal antibodies. *Chemical Engineering and Processing.* 44:1123-1137.

3. Shenoy, B., Wang, Y., Shan, W., and Margolin, L. 2000. Stability of crystalline proteins. *Biotechnology and Bioengineering.* 73:358-369

4. Weber, P. 1991. Physical principles of protein crystallization. *Advances in Protein Chemistry*. 41:1-36.

5. George, A., and W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallographica Section D.* 50:361–365.

6. Cantor, C., and Schimmel, R. 1980. Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function. W. H. Freeman and Company, San Francisco.

7. McMillan, W., and Mayer J. 1945. The statistical thermodynamics

of multicomponent systems. *Journal of Chemical Physics.* 13:276 –305.

8. Neal, B., Asthagiri, D., and Lenhoff, A. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophysical Journal.* 75:2469-2477.

9. Prasad, D., Virendra, U., and Rajiah, A. 1995. Dependence of the second virial coefficient on temperature. *The Chemical Engineering Journal and The Biochemical Engineering Journal.* 56:73-76

10. Vilker, V., Colton, C., and Smith, K. 1981. The osmotic pressure of concentrated protein solutions: effect of concentration and pH in saline solutions of bovine serum albumin. *Journal of Colloid Interface Science.* 79:548 –566.

11. Elcock, A., and McCammon, A. 2001. Calculation of weak protein-protein interactions: the pH dependence of the second virial coefficient. *Biophysical Journal.* 80: 613–625.

12. Dumetz, A., Snellinger-O'brien, A., Kaler, E., and Lenhoff, A. 2007. Patterns of protein-protein interactions in salt solutions and implications for protein crystallization. *Protein Science.* 16:1867-77.

13. Farnum, M., and Zukoski, C. 1999. Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophysical Journal.* 76:2716-2726.

14. Cantor, C., and Schimmel, R. 1980. Biophysical Chemistry: Part I: The conformation of biological macromolecules. W. H. Freeman and Company, San Francisco.

15. Rozhkov, S. 2006.Solubility and phase transitions in a water-protein-salt system. *Biofizika.* 51: 822-826.

16. Ruckenstien, E., and Shulgin, I. 2006. Effect of salts and organic additives on the solubility of proteins in aqueous solutions. *Advances in Colloid and Interface Science*. 123-126:97-103.

17. Cohn, E. 1925. The Physical Chemistry of the Proteins. *Physiological Reviews*. 5:349-437.

18. Roe, S. 2001. Protein Purification Techniques: A Practical Approach. Oxford University Press, New York.

19. Mcpherson, A. 2001. A comparison of salts for the crystallization of macromolecules. *Protein Science.* 10:418-422.

20. Kimber, M., Vallee, F., Houston, S., Necˇakov, A., Skarina, T., Evdokimova, E., Beasly, S., Christendat, D., Savchenco, A., Arrowsmith, C., Vedadi, M., Gerstein, M., and Edwards, A. 2003. Data-mining crystallization data-bases: knowledge-based approaches to optimize protein crystal screens. *Proteins: Structure, Function, and Genetics.* 51:562-568.

21. Lund, M., and Johnsson, B. 2003. A mesoscopic model for protein-protein interactions in solution. *Biophysical Journal.* 85:2940-2947.

22. Zhang, Y., and Cremer, P. Interactions between macromolecules and ions: the Hofmeister series. *Current Opinion in Chemical Biology.* 10:658-663.

23. Velev, O., Kaler, E., and Lenhoff, A. 1998. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophysical Journal.* 75:2682-2697.

24. Pjura, P., Lenhoff, A., Leonard, S., and Gittis, A. 2000. Protein crystallization by design: chymotrypsinogen without precipitants. *Journal of Molecular Biology.* 300:235-239.

25. Haas, C., Drenth, J., Wilson, W. 1999. Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution. *Journal of Physical Chemistry B.* 103:2808-2811.

26. Guo, B., Kao,S., McDonald, H., Asanov, A., Combs, L., and Wilson, W. Correlation of second virial coefficients and solubilities useful in protein crystal growth. *Journal of Crystal Growth.* 196:424-433.

27. Rupert, S., Sandler, S., and Lenhoff, A. 2001. Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnology Progress.* 17:182-187.

28. Liu, W., Bratko, D., Prausnitz, J., and Blanch, H. 2004. Effect of alcohols on aqueous lysozyme–lysozyme interactions from static light-scattering measurements. *Biophysical Chemistry.* 107:289-298.

29. Polson, A., Fotgeiter, G., Largier, J., Mears, G., and Joubert, F. 1964. The fractionation of protein mixtures by linear polymers of high molecular weight. Acta *Biochimica et Biophysica Acta.* 82:463-475.

30. Paliwal, A., Asthagiri, D., Abras, D., Lenhoff, A., and Paulaitis, M. 2005. Light-scattering studies of protein solutions:role of hydration in weak protein-protein interactions. *Biophysical Journal.* 89:1564-1573.

31. Arakawa, T., and Timasheff, S. 1985. Mechanism of poly(ethylene glycol) interactions with proteins. *Biochemistry.* 24:6756-6762.

32. Atha, D., and Ingham, K. 1981. Mechanism of precipitation of proteins by polyethylene glycol: analysis in terms of excluded volume. *The Journal of Biological Chemistry.* 256:12108-12117.

33. Ilett, S., Orrock, A., Poon, W., and Pussey, P. 1995. Phase behavior of a model colloid-polymer mixture. *Physical Review E.* 51: 1344-1352.

34. Tanaka, S., Ataka, M., Onuma, K., and Kubota, T. 2003. Rationalisation of membrane protein crystallization with polyethylene glycol using a simple depletion model. *Biophysical Journal.* 84: 3299–3306.

35. Kulkarni, A., Chatterjee, A., Schweizer, K., and Zukoski, C. 2000. Effects of polyethylene glycol on protein interactions. *Journal of Chemical Physics.* 113:9863-9873.

36. Baskir, J., Hatton, A., and Suter, U. 1986. Thermodynamics of the separation of biomaterials in two phase aqueous polymer systems: effect of the phase forming polymers. *Macromolecules.* 20:1300-1311.

37. Forcinit, D., and Hall, C. 1990. Theoretical treatment of aqueous 2-phase extraction by using virial expansions- a preliminary report. *ACS Symposium Series.* 419:53-70.

38. Abbott, N., Blankschtein, D., and Hatton, T. 1991. Protein partitioning in 2-phase aqueous polymer systems: novel physical pictures and a scaling thermodynamic function. *Macromolecules.* 24 : 4334-4348.

39. Abbott, N., Blankschtein., and Hatton, T. 1992. Protein partitioning in 2-phase aqueous polymer systems: a neutron scattering investigation of the polymer solution structure and protein-polymer interactions. *Macromolecules.* 25:3932-3941

40. Sheth, S., and Leckband, D. 1997. Measurements of attractive forces between proteins and end-grafted (poly)ethylene glycol chains. *Proceedings of The National Academy of Sciences of The United States of America. 94:8399-8404.*

41. Israelachvili, J. 1997. The different faces of (poly)ethylene glycol. *Proceedings of the National Academy of Sciences of The United States of America. 94:8378-8379.*

42. Bloustine, J., Virmani, T., Thurston, G., and Fraden, S. 2006. Light scattering and phase behavior of lysozyme-poly(ethylene glycol) mixtures. *Physical Review Letters.* 96:087803.

43. Lomakin, A., Asherie, N., and Benedek, G. 2003. Liquid-solid transition in nuclei of protein crystals. *Proceedings of The National Academy of Sciences of The United States of America.* 100:10254-10257.

CHAPTER 2

EXPERIMENTAL TECHNIQUES USED FOR MEASURING THE SECOND VIRIAL COEFFICIENT

As pointed out by Dumetz et al.[1], although measuring solubility seems to be a more apparent approach in characterizing protein interactions in the context of protein crystallization, solubility measurements depend on the formation of a solid phase, complicating the theoretical analyses as well as interpretation of the molecular mechanisms explaining the interactions on a thermodynamic basis. However, the second virial coefficient, as given by equation 1.2, depends only on the potential of mean force, and hence is the variable of choice when investigating weak protein-protein interactions and their thermodynamic origins.

This chapter will focus on a brief review of some of the analytical techniques used to measure B_{ii} values of proteins in solution. The main emphasis will be on static light scattering and its development from a batch method to a continuous flow-through method.

2.1 Membrane Osmometry

In chapter 1, equation 1.1 gave a definition for the second virial coefficient as a second order coefficient appearing in the expansion of solution osmotic pressure as a function of concentration. This definition is the underlying principle of membrane osmometry, one of the first and more traditional techniques used to measure the second virial coefficient[2,3,4]. The derivation of equation 1.1, as given by Cantor and Schimmel[5], is reviewed as follows:

Consider two chambers of equal volume, labeled 1 and 2, separated by a semi-permeable membrane, allowing only the passage of small solvent (water) molecules through it. Chamber 1 is filled with pure solvent, whereas chamber 2 contains a solution of solvent (water) + solute (protein) molecules. With the passage of time, solvent molecules will permeate from chamber 1 to chamber 2 (from higher to lower concentration), however, the large protein molecules will remain confined in chamber 2 due to their size. This retention of solute molecules is balanced by an increase in the solution pressure exerted by the contents of chamber 1. At equilibrium, this balancing pressure is known as the osmotic pressure and can be derived by equating the solvent chemical potentials on both sides of the membrane.

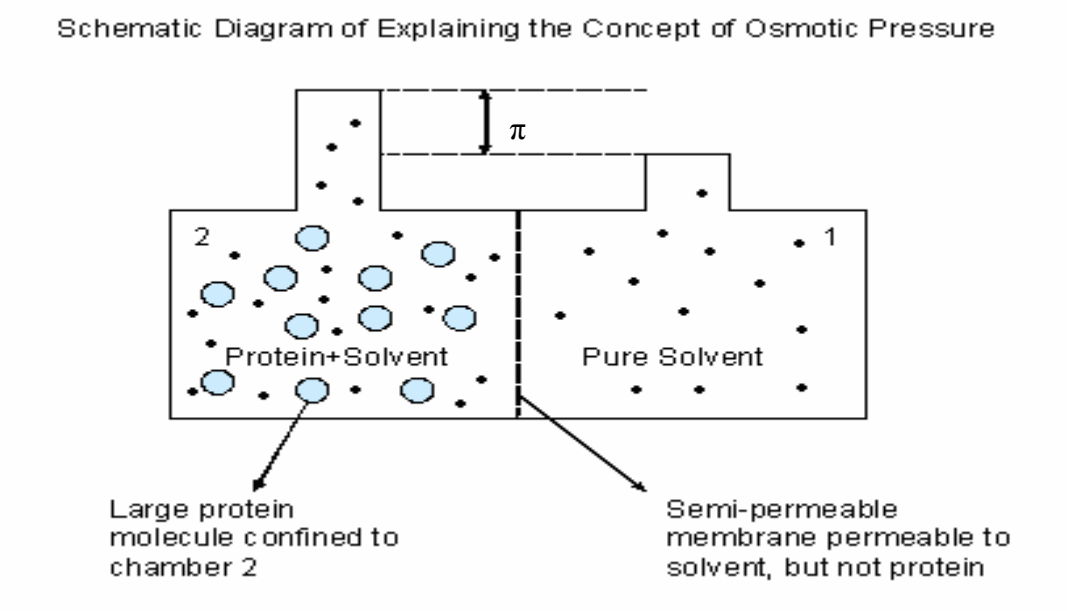

Figure 1. Schematic diagram explaning the concept of osmotic pressure.

Designating subscript s for solvent, i for protein, and notations 1 and 2 for the two chambers as described, the equality of solvent chemical potentials at equilibrium of chamber 1 and 2 is given as:

$$\mu_{s1}^{0}(T,P_1) + RT\ln \mathrm{X}_{s1} = \mu_{s2}^{0}(T,P_2) + RT\ln \mathrm{X}_{s2}, \qquad (2.1)$$

where, $\mu^{0}(T,P)$ is the standard chemical potential, and X is the mole fraction. X_{s1} is always unity as chamber one always contains pure solvent. Rearranging equation 2.1, we get,

$$\mu_{s1}^{0}(T,P_1) - \mu_{s2}^{0}(T,P_2) = -RT\ln \mathrm{X}_{s2} \qquad (2.2)$$

Assuming that the difference in standard chemical potentials of chamber 1 and 2 at equilibrium is solely due to pressure and taking a partial derivative of equation 2.2 with respect to pressure at constant temperature, we get,

$$\left[\frac{\partial \mu_{s1}^{0}(P_1,T)}{\partial P}\right]_T - \left[\frac{\partial \mu_{s2}^{0}(P_2,T)}{\partial P}\right]_T = RT\left[\frac{\partial \ln \mathrm{X}_{s2}}{\partial P}\right]_T \tag{2.3}$$

The left hand side of equation 2.3 is related to the partial molar volume, $\overline{\nu_s}$ and can be written as:

$$-\Pi\bar{\nu}_{s2} = RT\left[\frac{\partial \ln \mathrm{X}_{s2}}{\partial P}\right]_T, \tag{2.4}$$

where, $\Pi = (P_2 - P_1)$, the equilibrium osmotic pressure, and $\bar{\nu}_{s2} \approx M_s / c_{s2}$ for a dilute solution. Since $\mathrm{X}_{s2} = \dfrac{c_{i2}/M_i}{c_{s2}/M_s}$, equation 2.4 can be written as:

$$\Pi = RT\frac{c_{i2}}{M_i}, \tag{2.5}$$

which is the equation giving the osmotic pressure as a function of the solute concentration for ideal solutions. In real solutions, equation 2.5 is approached in a limiting form of concentration given by:

$$\Pi = \frac{RT}{M_i}c_{i2} + B_{ii}c_{i2}^2 + \ldots, \tag{2.6}$$

which can be compared to equation 1.1.

A membrane osmometer consists of a similar experimental setup as seen in figure 1., and by monitoring the change in osmotic pressure with the change in protein concentration for different protein samples, the second virial coefficient and the molecular weight of the solute can be obtained using equation 2.6. Solute concentration measurements are made externally, usually through UV spectroscopy for proteins.

This technique suffers many disadvantages, most of them associated with the membrane, that reduce the reliability and accuracy of measurements made through it. For proteins, most B_{ii} measurements are made in aqueous solvents, whereas the membranes are usually stored in organic solvents like methanol to prevent degradation[6].Thus, the experimental setup involves extensive conditioning to an aqueous solvent, which is both tedious and time-consuming. In addition, the membranes are extremely sensitive to temperature gradients, and can leak if subject to rapid expansion and compression due to fluctuations in temperature. Any leakage of solute to the pure solvent chamber will result in a lower osmotic pressure being recorded than the actual for the given solute concentration, and hence result in an underestimation of the B_{ii} value. Most commercial membrane matrices are made of cellulose acetate[6,7] that has a tendency to dry up and degrade very quickly if subjected to dry storage. At the same time, they are also water-soluble and hence offer limited lifetimes when used in aqueous systems.

The amount of data for second virial coefficients obtained using membrane osmometry is limited for protein solutions, which in part is explained by the various disadvantages associated with this technique.

2.2 Self-Interaction Chromatography

Self-interaction chromatography was developed as an alternative to osmometric and scattering (discussed later) methods of B_{ii} determination, aiming to minimize the amount the protein required as well as experimental run time[8]. The technique is based on the assumption that when a protein in solution partitions in a column covalently immobilized with the same protein, the average retention time is a measure of the interactions between the mobile and the immobile protein.

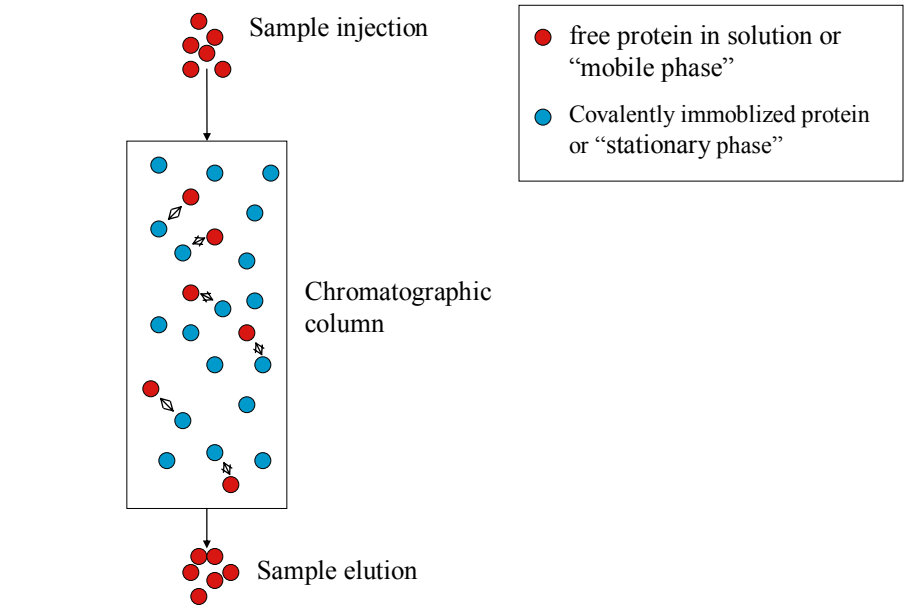

Figure 2. Schematic diagram explaining the concept of self interaction chromatography.

A brief description of the underlying theory relating the average retention time to the second virial coefficient is given as follows from Tessier et al[9].

Chromatographic retention time in a column is alternatively described in terms of the column retention factor, k'.

$$k' = \frac{V_r - V_o}{V_o}, \tag{2.7}$$

where V_r is the average retention volume, the mobile phase volume passed through the immobilized column from the time of injection to the time of elution of the solute, and V_o is the mobile phase required to elute a given solute through an "empty" column, i.e.,

without any immobilized protein. The column structure is assumed to be made up of an extra-particle or interstitial volume, the volume between immobilized protein molecules, and an intra-particle or pore volume, the volume within the immobilized molecules available for access to the mobile phase.

The retention factor is related to the free energy change, ΔG, in transferring a single solute particle of the mobile phase from the interstitial volume to the pore volume:

$$k' = \frac{\int_{V_p} \left(e^{-\Delta G / kT} - 1\right) dV_p}{V_o}, \tag{2.8}$$

where, V_p denotes the pore volume. This free energy change is assumed to be equivalent in definition to the potential mean of force evaluated at a fixed separation distance r_{12}, and angular orientations Ω_1 and Ω_2, where subscripts 1 and 2 denote the free and immobilized particles respectively.

$$\Delta G\left(r_{ij}, \Omega_i, \Omega_2\right) = W\left(r_{ij}, \Omega_i, \Omega_j\right) \tag{2.9}$$

The second virial coefficient is related to the potential mean of force according to equation 1.2 and can thus be calculated by measuring the retention factor via column chromatography. However, the equality assumed in equation 2.9 is a limited approximation, as by definition of the potential of mean force, the two interacting molecules should be free to sample all orientations in space, whereas, the immobilized molecule's orientation is fixed as it is covalently linked to the column. The theory is also limited in its assumption that the molecules in the mobile phase interact only with the immobilized molecules and not one another. In addition, it is assumed that at a given time, a single free protein molecule interacts with only a single immobilized protein

molecule, i.e, the interaction is assumed to be a "two-body" interaction. However, in a "tightly" packed column where the intermolecular spacing between immobilized particles is extremely small, it is inappropriate to model the interaction as two-body.

In addition to the inherent limitations of the theoretical model discussed above, the technique suffers from experimental setbacks as well that can complicate both data collection and analyses. There is a possibility of strong binding of impurities introduced from the buffers used to pack, wash, and elute the column (for a more detailed description of chromatographic protocols an reagents used, the reader is referred to Ahamed et. al[10]), which can affect the interpretation of B_{ii} values obtained through SIC. In addition, performance characteristics such as poor peak resolution (overlap between chromatographic peaks) can complicate the analysis of a chromatogram and hence the accuracy of the calculated virial coefficients as well.

SIC cuts down the experimental run time per virial coefficient measurement to approximately 45 minutes compared to osmometric and scattering methods that require several hours. However, experimental run-time alone is an inappropriate scale of comparison for "quickness" of data-collection as column preparation (treatment of "anchor" particles such as resins, sugars, etc. with suitable reagents before loading them into the column), packing (loading the column with the anchor particles followed by covalently linking protein molecules to the anchor particles), and equilibration (ensuring that all column parameters such as density of the packed particles, flow-rate of the immobile phase, etc. are steady) require at least two days of down time before any measurements can be made. While the long down-time may be offset by the consideration that once the column is immobilized, it can be used for up to two months;

using the same immobilized molecules for such a long duration generally denatures the protein as it is subject to multiple wash and eluting buffers repeatedly and necessitates column repacking. An elaborate description of the preparation, packing, and equilibration protocols employed in column chromatography can be found in the technical support literature published by chromatographic column manufacturers such as Pall Life Sciences, GE Healthcare, etc.

2.3 Static Light Scattering

Historically, the phenomenon of light scattering was first theoretically explained by Lord Rayleigh[11], where he discussed the scattering properties of particles much smaller (less than $1/20^{th}$) than the wavelength of the incident light. Lord Rayleigh's theory was extended by Debye[12] where he included the scattering effects of the shape of the molecules by accounting for wave interference between scattering from the same molecule. The Rayleigh-Debye theory was further extended by Zimm[13] for dilute polymer solutions, in which he accounted for the effect the size, or the "excluded volume" of the polymers on light scattering.

A "Zimm Plot" analysis is now a routine and standard way of characterizing macromolecules in solution using light scattering. A brief discussion of the development and construction of such a plot is given as follows:

Consider a sample volumeV, consisting of a dilute solution of single solute species in a given solvent, placed in the path of a monochromatic light beam of wavelength λ. As the light reaches the sample volume, most of it continues in the direction of its original propagation, however, a fraction of the incident light is scattered off of it in all directions

depending on the shape and size of the molecules comprising the sample volume. For a given solute of concentration c and molecular weight M_w, the intensity of the scattered light is directly proportional to the product of c, M_w, and $(dn/dc)^2$, where dn/dc is the differential refractive index for the given solute-solvent system, i.e. a measure of how the refractive index of the solution changes with a change in concentration of the solute.

$$I \propto cM_w(dn/dc)^2 \tag{2.10}$$

The above approximation is further modified to include the second virial coefficient to account for the effect of solute-solute interactions on scattering intensity mediated by the solvent. Physcically, this can be conceptualized as particles in close proximity of one another (and interacting with each other as a result of their proximity) scattering light differently than as individual molecules in solution.

Assuming that the solute particles are less than $1/20^{th}$ the size of the incident wavelength (also known as the Rayleigh-Gans-Debye limit[14]) and that this size is so small that any angular variations in light scattering can be ignored, the following equation was obtained by Zimm :

$$\frac{Kc}{R(\theta)} = \frac{1}{M_w} + 2B_{ii}c \tag{2.11}$$

$R(\theta)$ is the excess Rayleigh ratio, the ratio of scattering intensity of the solution to the scattering intensity of the pure solvent. K is an optical constant determined by the wavelength of the incident light and properties of the solution:

$$K = \frac{4\pi^2[n_s(dn/dc)]^2}{N_A\lambda^4} \tag{2.12}$$

where, n is the refractive index of the solution, n_s is the refractive index of the solvent, and N_A is Avogadro's number.

A Zimm plot is obtained by measuring and plotting the excess Rayleigh ratio as a function of solute concentration according to equation 2.12, from which several important physical properties of the solute can be determined, namely, the molecular weight, M_w, and the second virial coefficient, B_{ii}.

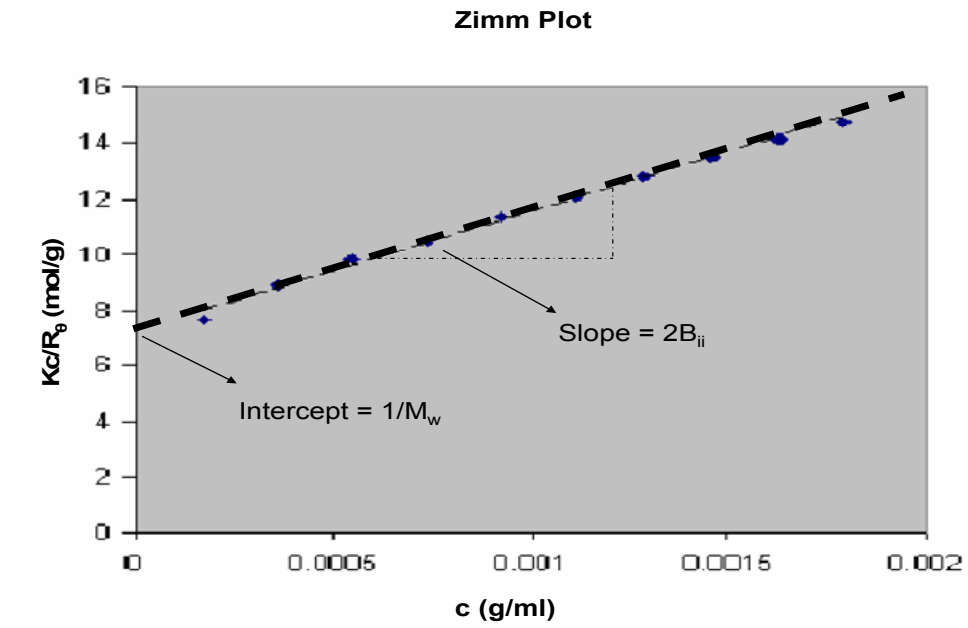

Figure 3. Example of a Zimm Plot

This treatment assumes that the intensity of light scattered is uniform at all angles from the incident light due to the small size of the particle, however, angular variation in light scattering also allows for the determination of the radius of gyration, R_g, of the particle[15].

The instrumentation of a modern light scattering system is discussed briefly in the next paragraph as it is the chosen technique to measure virial coefficients in this study.

There are three main components to a light scattering system, namely, the light source, the sample cell, and the detector itself. Traditionally, before the invention of lasers, the light source was usually a mercury lamp placed with optical filters to obtain a monochromatic light beam. The first laser-based light scattering system was developed by Wyatt[15], in 1971. The sample cell is either a static cell such as a cuvette, or a flow-cell designed for continuous operation. Earlier systems employed a detector called a "scanning goniometer", a single detector that moved about an arc around the sample cell. Although "multi-angle" in its sense of rotation, it still was a fixed-angle design as scattering was measured from only one given angle at a given time. This scanning detector design was replaced by multi-angle detector layout. For e.g. most of the DAWN series of Wyatt Technology's detectors are based on a multi-angle photometer design where multiple detectors (3 to 18) are placed at specific angles with respect the light beam around the sample cell. The multi-angle design is important to account for angular effects observed with large particle sizes, however, under the Rayleigh-Gans-Debye limit, scattering data is usually obtained at a single angle of 90^{o}. The 90^{o} placement is chosen as higher and lower angles are prone to heavy signal noise due to scattering from dust particles and other contaminants that may enter the flow-cell. The concentration measurements are made externally using UV spectrophotometry or differential refractometry.

Most light scattering systems are designed to operate in a batch mode, where scattering measurements are made for a given solute concentration contained in a cuvette.

For a Zimm analysis, which calls for multiple concentration measurements, the method involves making several stock solutions and obtaining individual readings for each concentration by manually replacing the sample in the cuvette. This is not only time consuming, but also requires a lot more solute, which can be a limitation with proteins. Alternatively, a flow-cell design containing the sample volume allows for continuous measurement of the scattering intensities as the sample at a given concentration passes through at a given flow rate. The continuous method is advantageous over the batch mode of measurement in terms of amount of run time to obtain a single virial coefficient measurement, although it doesn't necessarily save on the amount of sample used as it still requires multiple stock solutions at varying concentrations.

More recently, Minton[16] designed a "composition gradient" pumping mechanism in order to study reversible hetero-associations between proteins. The composition gradient method delivers a continuous step-wise gradient of solute concentrations to the flow-cell from a single stock solution. This method minimizes the amount of experimental run time as well as the amount of sample being used. The design was adapted and commercialized by Wyatt Technology to obtain virial coefficient, molecular weight, and radius of gyration measurements in addition to hetero-association data. The composition gradient method will be discussed in more detail in the following section of this chapter. Since light scattering data is greatly affected by foreign scatterers like dust particles, extreme care has to be taken during sample preparation and cuvette/flow-cell cleaning to eliminate contamination. Most fundamental texts on light scattering[17] from dilute solutions recommend a tedious sample dialysis (several days in large volumes of buffer) to ensure equilibrium. In addition to the long dialysis times, dialyzing protein samples

extensively against buffers can result in sample aggregation as well as loss on the dialysis membrane As a result, dialysis is often overlooked and omitted from the sample preparation methods, but can significantly improve the data quality and reliability if done properly.

2.4 Composition Gradient Static Light Scattering

The composition gradient method is an innovative sample delivery method developed by that uses programmable syringe pumps to generate a step-wise concentration increment or decrement from an initial stock concentration, to be used upstream to the light scattering and concentration detectors. The analysis remains the same as with batch light scattering and is governed by equations 2.11, and 2.12. Figure 4 gives a schematic diagram of the plumbing in a composition gradient assembly.

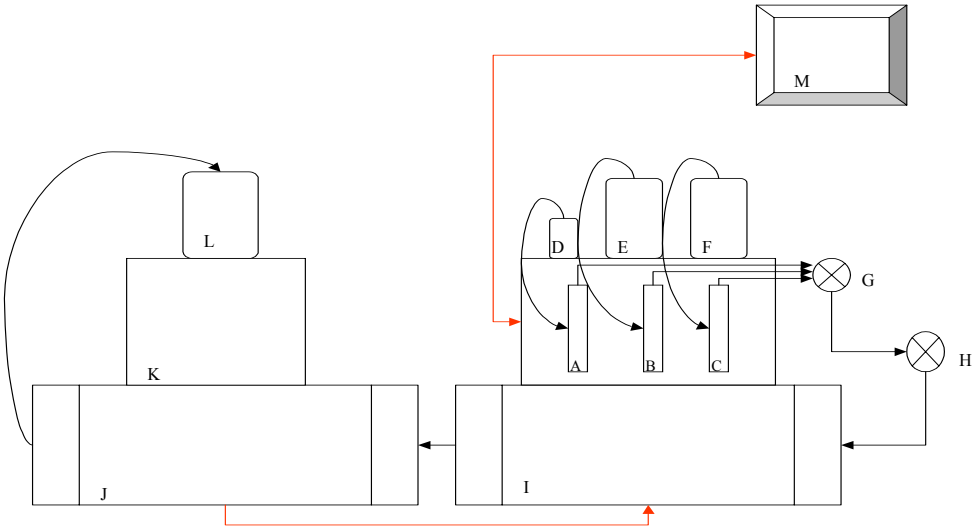

Figure 4: Schematic diagram of a composition gradient assembly-A,B,C: programmable syringe pumps; D,E,F: solute, solvent, and wash reservoirs respectively; G: mixer; H: in-line filter; I: static light scattering detector; J: concentration detector; K: Stand; L: waste reservoir; M: computer workstation; black arrows indicate flow lines and red arrows indicate electronic data transmission cable

By using a computer to control the flow-rates, a combination of three syringe pumps can be programmed to deliver a step-wise concentration gradient to a light scattering and concentration detector in series.

Figure 5 illustrates an example of a gradient design used to obtain a Zimm plot in one single continuous experimental run.

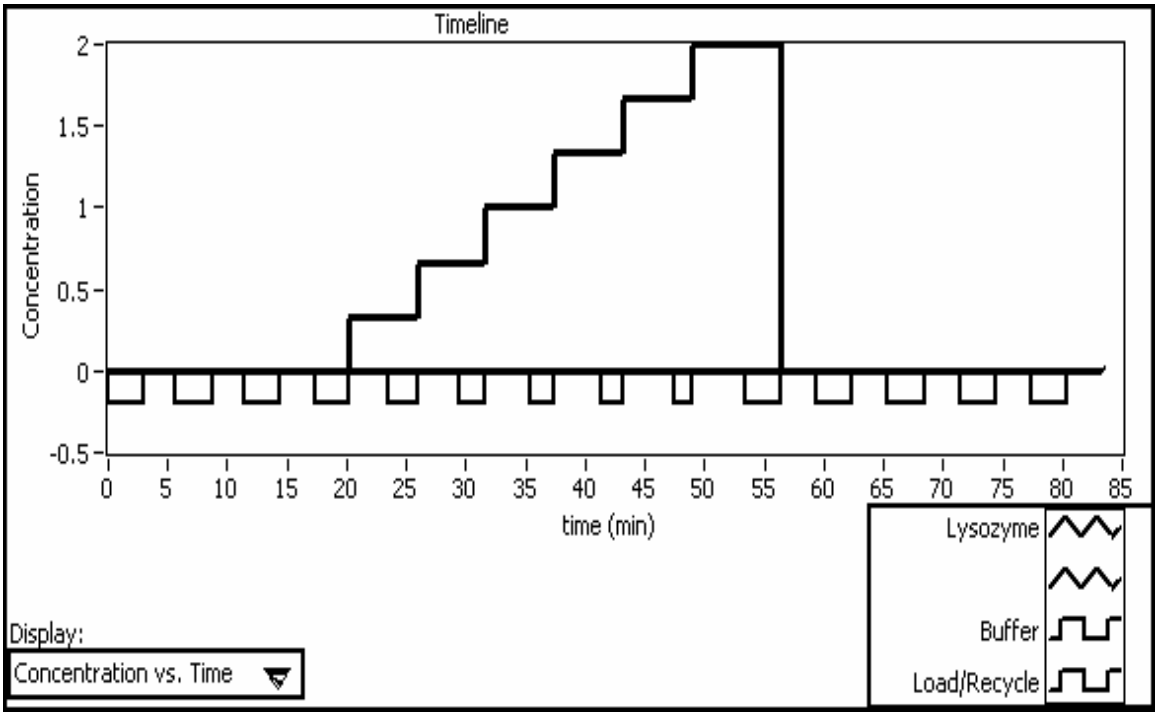

Figure 5. Illustration of a composition gradient run

For example, the gradient in figure #5 is created by programming the syringes such that the syringe connected to the lysozyme reservoir is increasing its flow-rate in incremental steps, whereas the syringe connected to the diluting buffer is simultaneously decreasing its flow-rate. The third pump remains idle in this design. The net result is a decrease in lysozyme dilution over six steps, with each step corresponding to a certain concentration of lysozyme.

Although Minton's original design had the two detectors in parallel, extra care has to be taken to ensure that the flow-rate in both parallel channels is equal. This method uses only a few ml of dilute sample solution to measure a B_{ii} value in less than 30 minutes. Thus, it offers the ease of light scattering analyses at similar time-scales to faster

techniques like SIC. The in-line filter serves as an additional mixer for the generated composition to ensure a clean and well-mixed flow to the detectors. In collaboration with Minton, Wyatt Technology has designed a beta version of the composition gradient system, CALYPSO™.

2.5 References

1. Dumetz, A., Snellinger-O'Brien A., Kaler, E.W., and Lenhoff, A. 2007. Patterns of protein protein interactions in salt solutions and implications for protein crystallization. *Protein Science.* 16:1867-1877

2. Shaink, H., and Smit, J. 2000. Determination of the osmotic second virial coefficient and the dimerization of β-lactoglobulin in aqueous solutions with added salt at the iso-electric point. 2000. *Physical Chemistry, Chemical Physics.* 2:1537-1541.

3. Reˇsˇciˇc, J., Vlachy, V., Jamnik, A., and Glatter, O. 2001. Osmotic pressure, small angle x-ray, and dynamic light scattering studies of human serum albumin in aqueous solutions. *Journal of Colloid and Interface Science.* 239:49-57.

4. Haynes, C., Tamura, K., Korfer, H., Blanch, H., and Prausnitz, J. 1992. Thermodynamic properties of aqueous α-chymotrypsin solutions from membrane osmometry measurements. *Journal of Physical Chemistry.* 96:905-912.

5. Cantor, C., and Schimmel, R. 1980. Biophysical Chemistry: Part I: The conformation of biological macromolecules. W. H. Freeman and Company, San Francisco.

6. Schafer, O. 1979. Modern aspects of membrane osmometer design. *Colloid and Polymer Science.* 257: 1007-1008.

7. Birdi, K. 2005. Membrane osmometry of aqueous micellar systems. *Colloid and Polymer Science.* 250:731-735.

8. Tessier, P., Lenhoff, A., and Sandler, S. 2002a. Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. *Biophysical Journal.* 82: 1620–1631.

9. Tessier, P., Sandler, S., and Lenhoff, A. 2004. Direct measurements of protein osmotic second virial coefficients by cross interaction chromatography.

10. Tangir, A., Ottens, M., Gijs, W., van Dedem, and Luuk., A. 2005. Design of self interaction chromatography as an analytical tool for predicting protein phase behavior. Journal of Chromatography A. 1089:111-124.

11. Lord Rayleigh. 1871. On the light from the sky, its polarization and color. *Philosophical Maganzine.* 41:107-120;274-279.

12. Debye, P. 1944. Light scattering by concentrated polymer solutions. *Journal of Applied Physics.* 15:338.

13. Zimm, B., 1948. The scattering of light and the radial distribution function of high polymer solutions. *Journal of Chemical Physics*. 16:1093.

14. Lyklema, J. 1991. Fundamentals of Interface and Colloid Science. Morgan Kauffman, San Francisco.

15. Wyatt., P. 1993. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta.* 272:1-40.

16. Kameyama, K., and Minton, A. 2006. Rapid quantitative characterization of protein interactions by composition gradient static light scattering. *Biophysical Journal.* 90:2164-2169.

17. Brown, W. 1996. Light scattering- principles and development. Oxford University Press, New York.

CHAPTER 3

OBJECTIVES

3.1 Characterization of Composition Gradient Static Light Scattering

The first objective of this thesis is to investigate the suitability of the composition gradient method as a technique for rapid and inexpensive measurements of protein interactions in solution. The approach used for process characterization is based upon systematic identification and screening of process variables specific to the composition gradient system, and comparison of obtained data after optimizing the operating conditions with published sets of data.

After examining the system, the following process variables were identified: detector flow-rate, number of concentration plateaus, initial protein stock concentration, and the type of in-line filter. These variables were first examined independent of each other, and then combined into factorial experiments to screen their effects.

3.2 Investigation of Polyethylene Glycol on Protein Interactions in Solution through Composition Gradient Method

The second objective of this work is to investigate the effect of PEG (molecular weight = 400 g/mol) concentration on lysozyme interactions in solution and compare the results obtained from those prediced by interactions on account of a depletion effect alone.

CHAPTER 4.

EXPERIMENTS

4.1 Process Variables

A brief description of all the process variables examined for the characterization of the composition gradient method is given in this section.

4.1.1 Flow-rate

There are three different flow-rates associated with the composition gradient system, namely, the syringe pickup rate, syringe dispense rate, and the detector flow-rate. The syringe pick-up rate is the rate at which the syringes draw fluid from the respective reservoirs, through a membrane degasser. Thus, the pick-up rate directly determines the degassing efficiency. It is important to allow sufficient degassing of the fluid as any gas/air entering the system may outgas in the detectors giving rise to false signals. The dispense rate is the rate at which the syringes dispense fluid to the detectors connected downstream to the composition gradient system. The detector flow-rate is the sum of all the individual dispense rates from all the syringes and thus is the "total" flow-rate in the detectors downstream. Higher detector flow-rates result in better mixing, however, they also generate high system back-pressure (upstream pressure exerted due to downstream components such as tubing, connectors, etc. determined by the geometry of the

components as well as parameters of the fluid such as flow-rate, viscosity etc.), which can damage the syringes.

The detector flow-rate is an important variable requiring optimization, where an optimal flow-rate is defined as one that offers uniform mixing, minimal signal noise, and minimal back-pressure.

4.1.2 Number of Concentration Steps

The combination of three syringes is used to generate a concentration gradient with time. The number of concentration steps, or plateaus, corresponds to the number of data points used in the construction of a Zimm Plot, and hence directly affects the reliability of the plot. It is important to characterize the optimum number of steps required to give consistent data while conserving the amount of sample. The following figure is an example of a step gradient design, where the syringes are programmed to deliver six equal incremental concentration steps of lysozyme from 0 to 2 mg/ml.

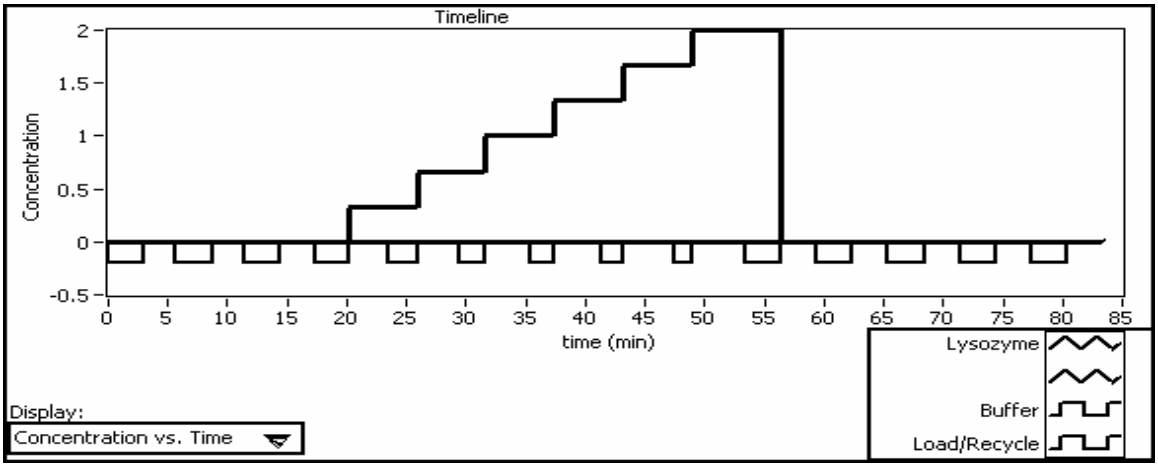

Figure 6. Illustration of a 6 step composition gradient run

The number of concentration steps not only determines the number of data points in a Zimm plot, it also determines the "jump" between each data point. Fewer steps in the gradients would mean fewer data points spaced farther apart in magnitude, whereas, a large number of steps would mean more data points spaced closer together.

It is important to determine the effect of the number of concentration steps on the detector sensitivity to ensure the reliability of the gradient design.

4.1.3 Initial Sample Stock Concentration

The concentration gradient is generated from a sample solution of fixed stock concentration by diluting with the required amount of buffer in line. Thus, the stock concentration determines the maximum value of protein concentration that can be measured, or the highest concentration data point in a Zimm Plot. For example, if the initial stock concentration is 5 mg/ml protein and the gradient is designed to deliver 10 concentration steps, each step will consist of concentrations in the increments of 10% of the 5 mg/ml, i.e. .5 mg/ml, 1 mg/ml, 1.5 mg/ml etc.

The sample stock concentration is an important process variable as it is directly related to the sensitivity of both the light scattering as well concentration detectors. It also determines the amount of protein required for an experimental run along with the number of concentration steps.

4.1.4 Type of In-line Filter

After the syringes deliver the required composition to the mixer, the mixed fluid passes through an in-line filter before going to the detectors. The filter serves two purposes, namely, to remove any particulate debris that may have entered the flow due to handling

or a dirty flow-cell, as well as to act as a secondary static mixer to ensure the concentration is well mixed in order to get a uniform signal.

The composition gradient assembly from Wyatt Technology is supplied with a filter-housing from Upchurch Scientific capable of carrying a 13mm filter disc. The choice of the filter matrix is governed by its compatibility with the solvent and sample species as well as their sizes. For this system, the filter membrane should be hydrophilic as the buffers are aqueous. In addition, hydrophobic surfaces can trigger denaturation of the protein sample. The filter should also have minimal affinity for the protein to minimize sample loss via adsorption. Finally, the porosity of the filter should be large enough to allow all species of interest to pass through and reach the detectors, yet small enough to screen possible contaminants. Keeping these characteristics in mind, suitable membrane matrices, namely cellulose acetate, PVDF (polyvinylidene difluoride), and PES (polyethersulfone), were identified. The membrane porosity was fixed at .22μm as it offers tight control over micro-contaminants while allowing nanometer sized proteins through.

Millipore's durapore PVDF matrix has the smallest protein binding coefficient (measured in terms of micro-gram protein adsorbed per cm^2 of the matrix) of all the materials identified and was finalized as the filter matrix to be used in this study.

4.2 Screening Experiments

4.2.1 Experimental Design

A simple 2^3 factorial design (3 variables studied at two levels each in one single experiment) was the starting point to screen the effect of the detector flow-rate, number of concentration steps, and the initial stock concentration on a response output, the molecular weight of the protein. Each of the variables was designated a "high" and a "low" operating level as described:

The recommended limit for the maximum operating detector flow-rate is 1 ml/min when the light scattering and concentration detectors are connected in series. A "high" of .5 ml/min was established for the experiment to stay well below the maximum limit, whereas a "low" was set to .2 ml/min. A "high" of 10, and a "low" of 5 was set for the number of concentration steps. From the data in the literature, a typical batch experiment consists of concentration measurements between 1 and 10 mg/ml protein. However, this system employs a UV detector to make protein concentration measurements and hence the stock concentration is limited by a value determined by the extinction coefficient of the protein, and the path-length of the UV flow-cell. For lysozyme, the extinction coefficient is about 26.5 cm^{-1} and the path-length is 4 mm, thus, according to Beer-Lambert's law, the UV detector can only detect lysozyme concentrations upto approximately 4 mg/ml. Thus, a "high" of 3.75 mg/ml and a "low" of 1 mg/ml was set for the initial stock concentration. The experimental design is summarized in table 1 given in the appendix.

4.2.2 Screening Results

The following data sets are a comparison of light scattering signals from run # 1 with low detector flow-rate and run # 2 with high detector flow-rate.

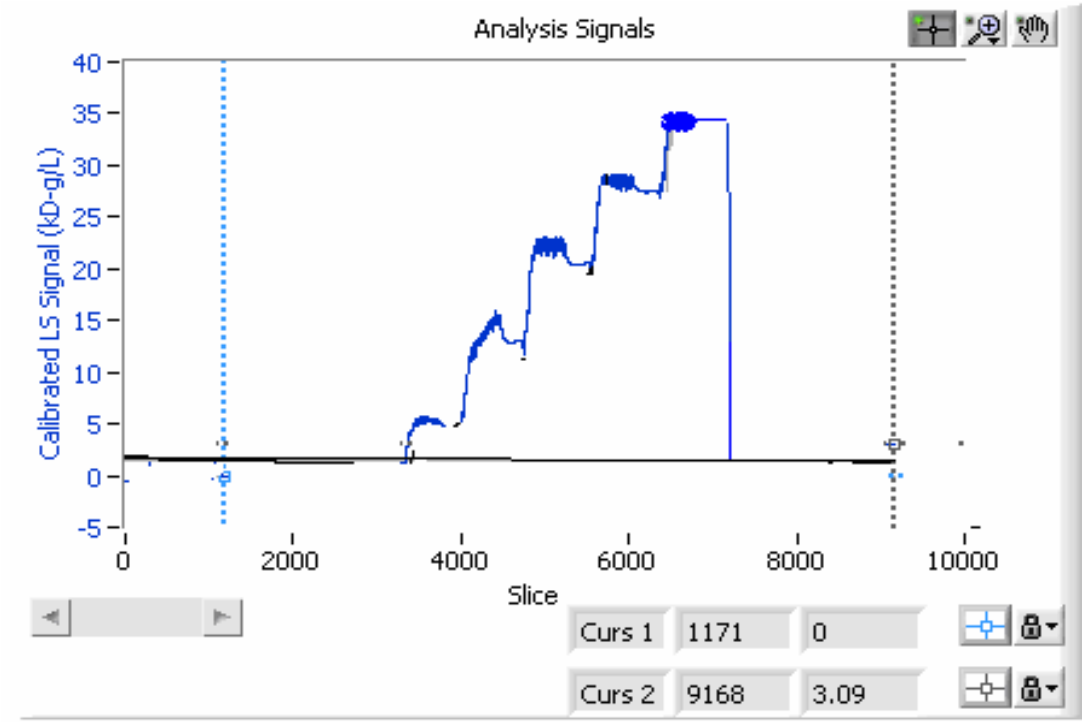

Figure 7. Illustration of noisy data obtained at low detector flow rates

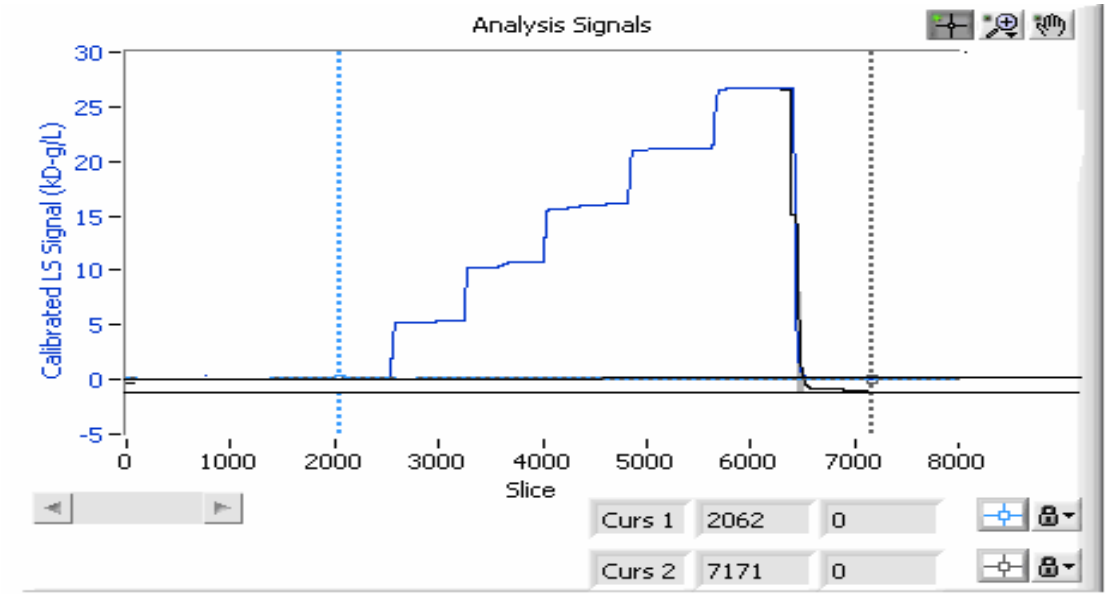

Figure 8. Illustration of noise-free data obtained at high detector flow rates

Signals from the first run are not only noisier than from the second, there is also a considerable "relaxation" of signals in each step of the first run. This relaxation is completely absent from the second run indicating that higher flow-rates give better mixing of the compositions generated by the pumps. The relaxation was consistently absent from all the runs with a high flow-rate except where the number of concentration steps was high as well.

The following data sets compare the light scattering signals from run # 2 with 5 concentration steps (low) spaced farther apart and run # 5 with 10 concentration steps (high) spaced closely. The detector flow-rates for both runs was high.

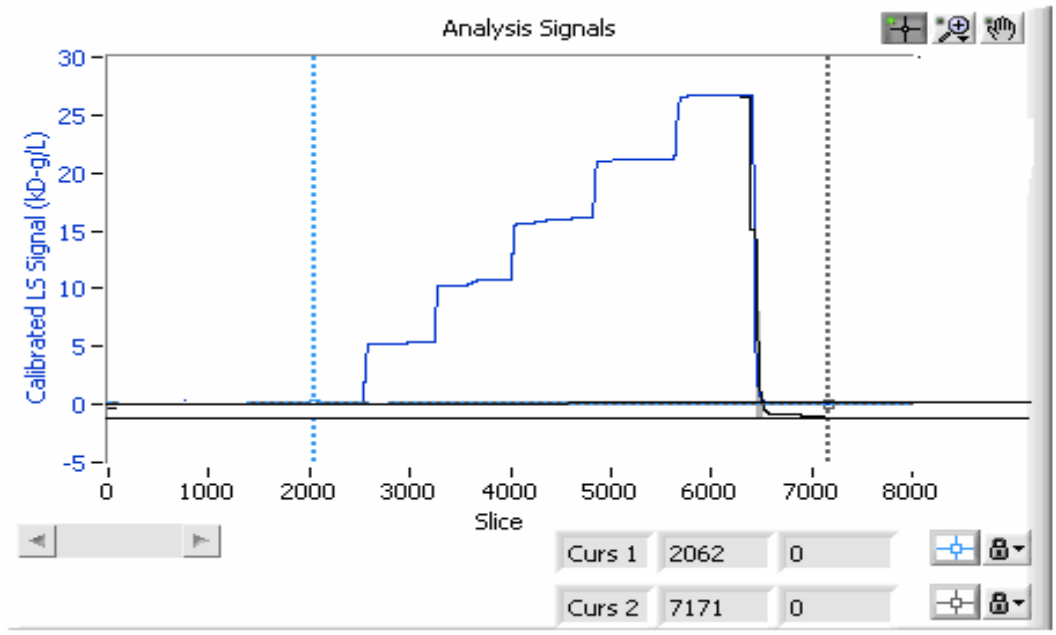

Figure 9. Illustration of noise-free data obtained with smaller gradient size

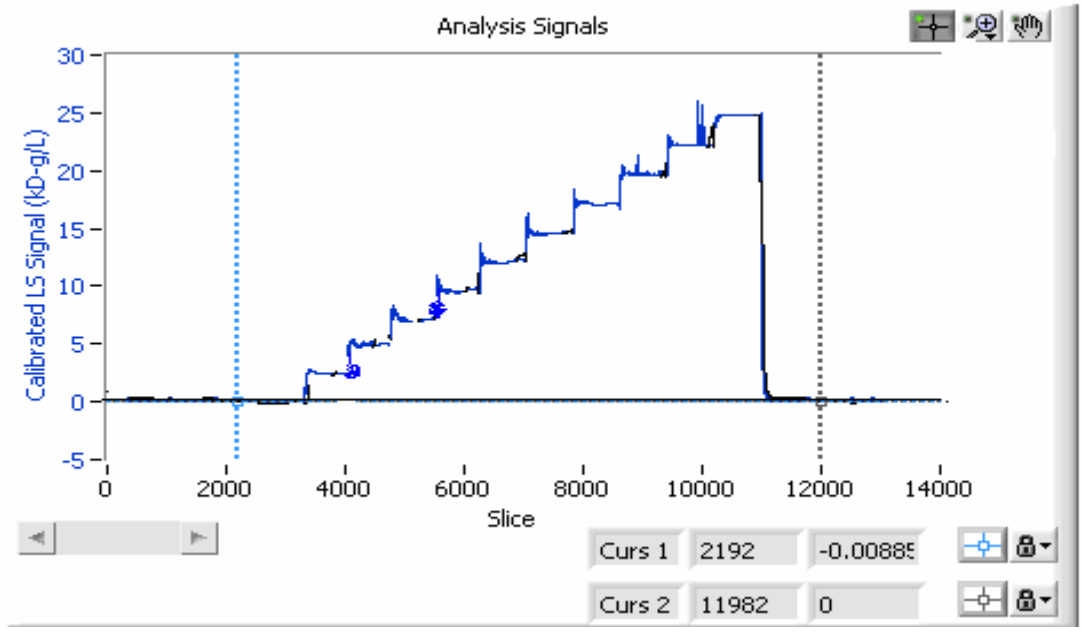

Figure 10. Illustration of noise introduced with large number of steps in gradient

It was observed that widely spaced concentration values improved the signal to noise ratio compared to closely spaced concentrations. However, this effect was not independent of the detector flow-rate and was more pronounced at high flow-rates.

It was observed that high initial stock concentration caused the UV detector signals to "max" out after three steps of the gradient irrespective of the number of steps remaining. No distinct effect of the initial stock concentration on the light scattering signals was observed. The following data set from run # 8 illustrates the "maxing out" effect observed in the UV signals at high initial stock concentrations.

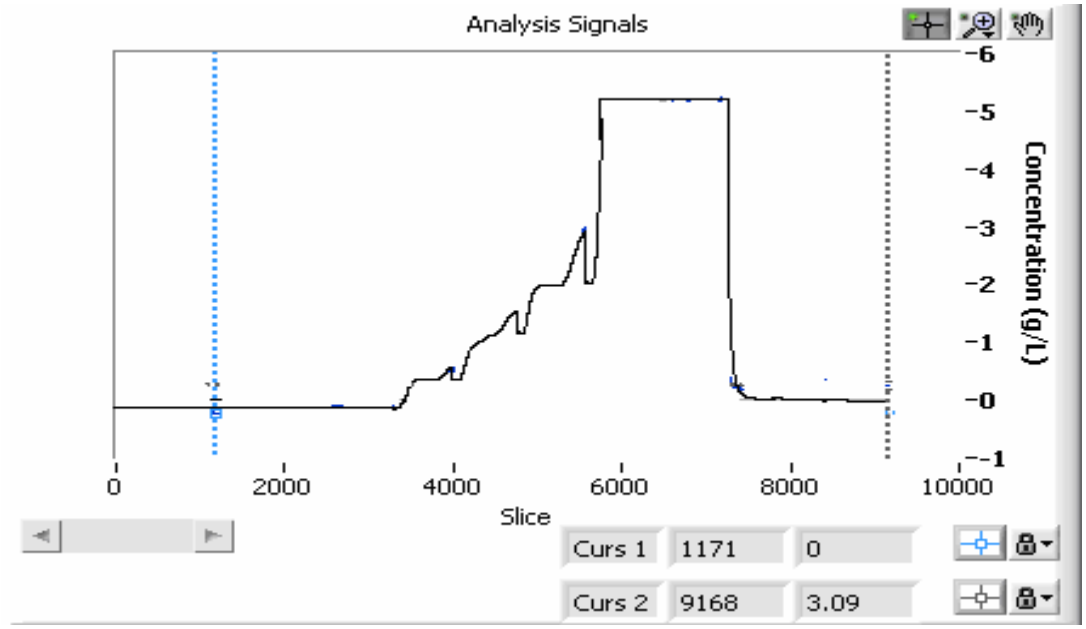

Figure 11. Illustration of "maxing" out of UV signals when high initial stock concentrations are used

In addition to the signal attenuation, a "drift" in the signal is observed at the end of every injection. During this time, there is no flow through the detectors due to the syringes loading for the next injection.

4.2.3 Conclusions and Recommendations

Based on the screening results, the operating detector flow-rate should be close to .5 ml/min. The number of steps in a gradient should be below 10 – if required, a run with fewer concentration steps should be repeated twice for accuracy as opposed to a single run with a large number of steps. The extinction coefficient of the protein decides the maximum concentration that can be determined from the UV signals.

Based on the tendency of the UV detector to cause signal attenuation when higher stock concentrations are used, the initial stock concentration should be less than 50% of the maximum theoretical value that can be determined. In addition, the UV flow-cell should be checked to see if it is fitted properly. The drift in the UV signal observed at the end of every injection could be a result of a loss in system pressure due to a loosely fitted cell.

4.3 Improvised Experiments Based on Screening Results

4.3.1 Experimental Design and Results

An eight step gradient was designed with the detector flow-rate set to .33 ml/min. The aim of the experiment was to observe the quality of the data with the optimized parameters while comparing the value of the molecular weights as well as virial coefficients obtained with published values in the literature. The results are summarized in Table 2.

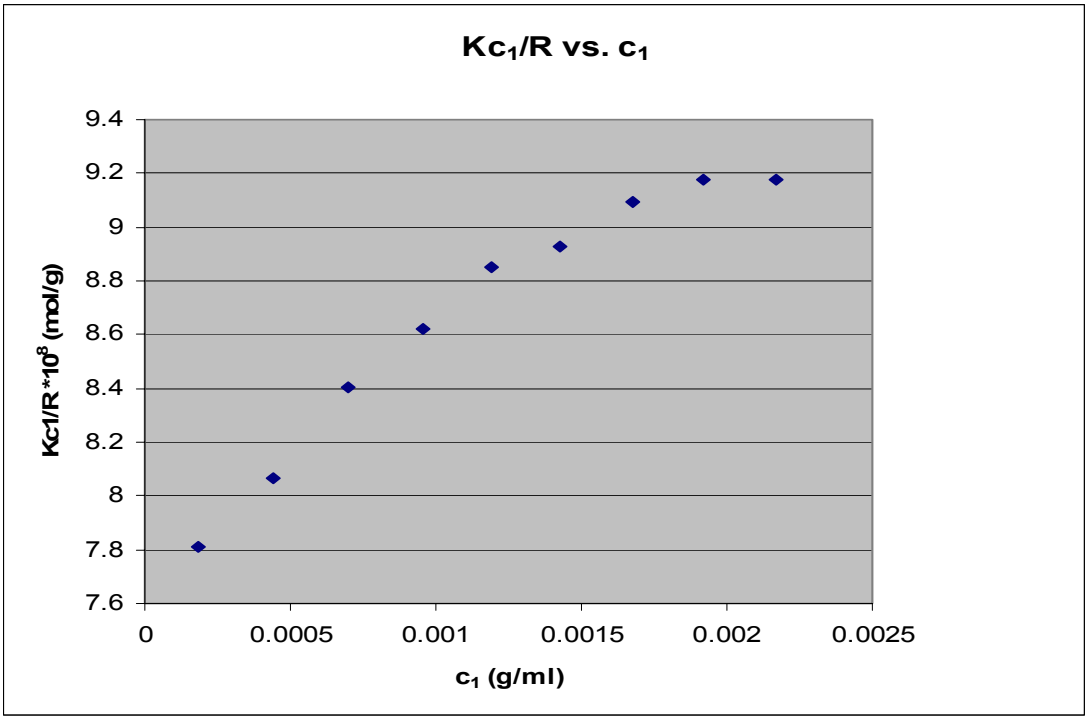

Figure 12. Zimm Plot of data obtained from run with improvised screening parameters; M_1 = 50000 g/mol; $B_{11} = 3.5*10^{-6}$ mol.ml/g^2

As given by the manufacturers, the molecular weight for a lysozyme monomer is only 14600 g/mol. Although no phase separation was visible to the naked eye, the light scattering data suggests that the lysozyme in the stock solution is not present as a monomer but a trimer (~50/14). The reason for trimerization of the protein could possibly be due to two reasons: variability in the buffering action (pH control) of Tris buffer as a result of a lack of ambient and system temperature control, and/or, repeated freezing/thawing of stock protein. The lysozyme used for making stock solutions is stored at -20^0 C, and needs to be thawed everytime before making the samples. Proteins are known to aggregate when subjected to repeated freeze/thaw cycles during storage.

Since the scattering indicates that proteins are not in their monomeric form, the self virial coefficient obtained reflects the interactions between two such trimers. The value of B_{11} obtained from this run is positive but extremely small and outside the "crystallization slot",-8 to -1 mol.ml/g^2, as reported by George and Wilson[1], indicating a weak repulsion between two protein trimer molecules of ~ 50 kD each.

In addition to the trimerization of the sample, the data also indicates a "plateau" in scattering as a function of concentration at higher protein concentrations. The plateau effect was repeatedly observed in subsequent runs of the same experiment and its onset seemed to be concentration dependent, i.e, a flattening of the slope of the Zimm plot was apparent at protein concentrations as low as > 1mg/ml and as high as 2 mg/ml.

4.3.2 Conclusions and Recommendations

Investing in a temperature controlled detector will help eliminate any changes in sample conformation due to temperature fluctuations during the experiment. Ordering only the required amount for protein for an experiment just ahead of the planned runs instead of purchasing proteins in "bulk" amount and freezing them for repeated use will eliminate any changes occurring in the sample conformation due to prolonged storage. While it is convenient to bulk order lysozyme as it is relatively cheap (and often available commercially only in bulk quantities), doing so requires freezing/thawing the entire lot to measure out the required amount for each experiment. Another solution would be to ration the bulk ordered protein into individual lots based on a "typical" experimental design as soon as it is received, and then freezing- this way, only the required amount of protein is thawed at a given time.

The flattening of the slope at concentrations above 1 mg/ml does not make physical sense since a flat slope theoretically indicates the absence of interactions between solute molecules and approximates ideal behavior. To determine if the plateau effect is introduced due to the system instrumentation, the gradient design should check for hysterises.

4.4 Hysterises Check

4.4.1 Experimental Design and Results

Four different gradients were designed to check for hysterises effets. The gradients and the results obtained are discussed in the following section.

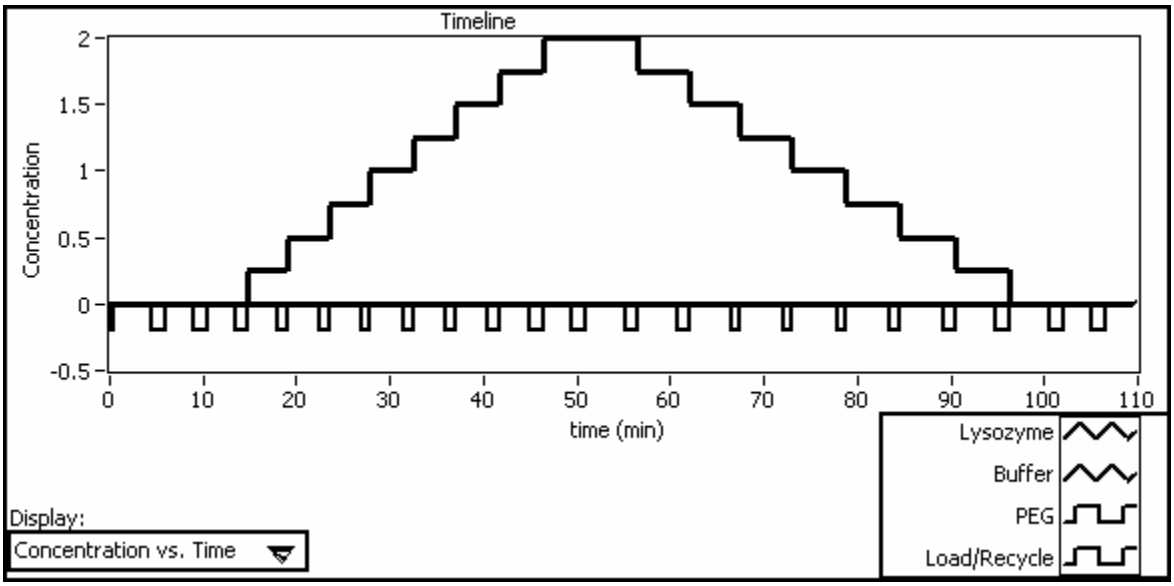

Figure 13. Illustration of gradient design #1 of hysterises experiments

The first half of the run is an ascending gradient where the concentration of lysozyme increases in increments of 12.5% with each step. After reaching 100% of the stock concentration, the concentration of lysozyme then decreases in decrements of 12.5% during the rest of the run. Theoretically, assuming the data collection is free from hysterises, the data points from the ascending gradient should overlap with the data points from the descending gradient. The results obtained are summarized below:

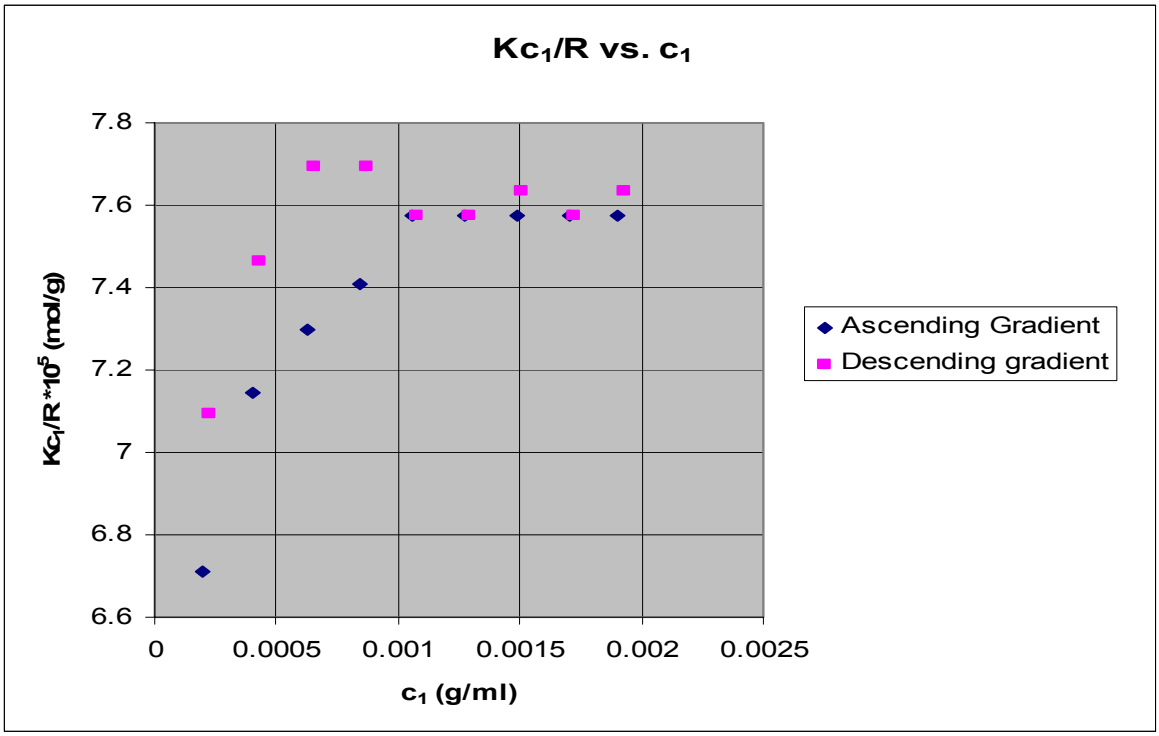

Figure 14. Zimm plots obtained from gradient #1 of hysterises experiments; M1 = 14430 (ascending gradient) ; M1 = 13605 (descending gradient)

From the data, it is clear that there is a good overlap in the corresponding x axis (concentration) values for the two gradients, indicating that the compositions generated by the pumps are independent of the order in which they are generated.

The scattering data for corresponding concentration points from the ascending and descending gradients seem to overlap at higher concentrations (>1 mg/ml) but differ greatly for lower concentrations, however, the agreement of the data points cannot be attributed to concentration since a plateau is once again above concentrations of 1 mg/ml. Fitting data points below 1 mg/ml from both the ascending and descending gradients, the following molecular weights were obtained:

According to the manufacturer, $M_1 = 14600$, thus the error margin in the ascending gradient is -1.17% and -6.81% for the descending gradient.

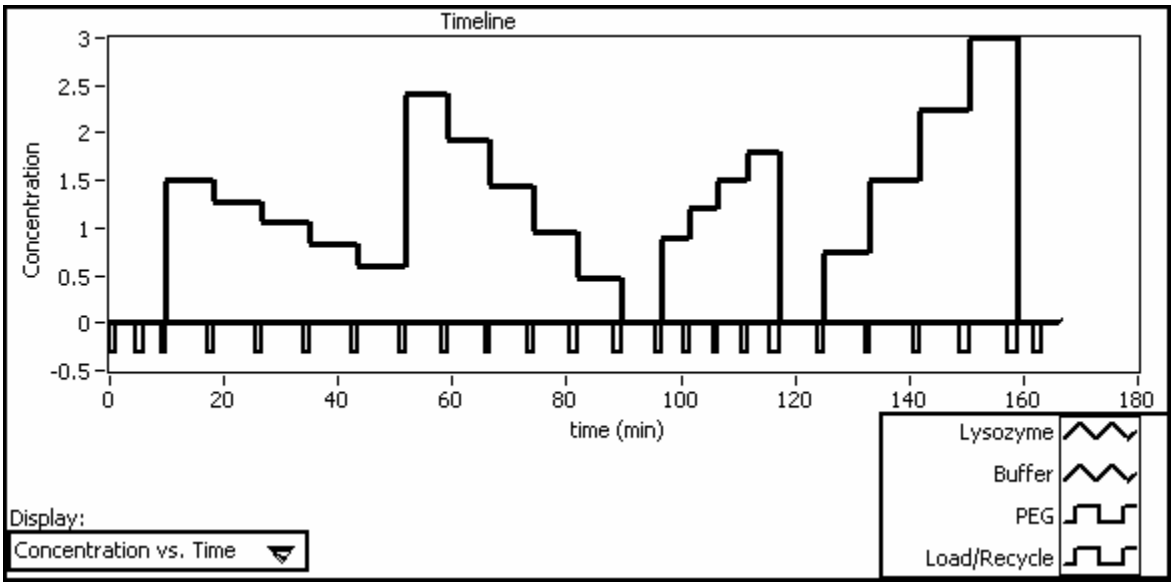

Figure 15. Illustration of gradient design #2 of hysterises experiments

The second gradient design used to test for hysterises effects is illustrated above. It consists of four randomly arranged smaller descending and ascending gradients, where the magnitude of the concentration step is different for each of the gradients.

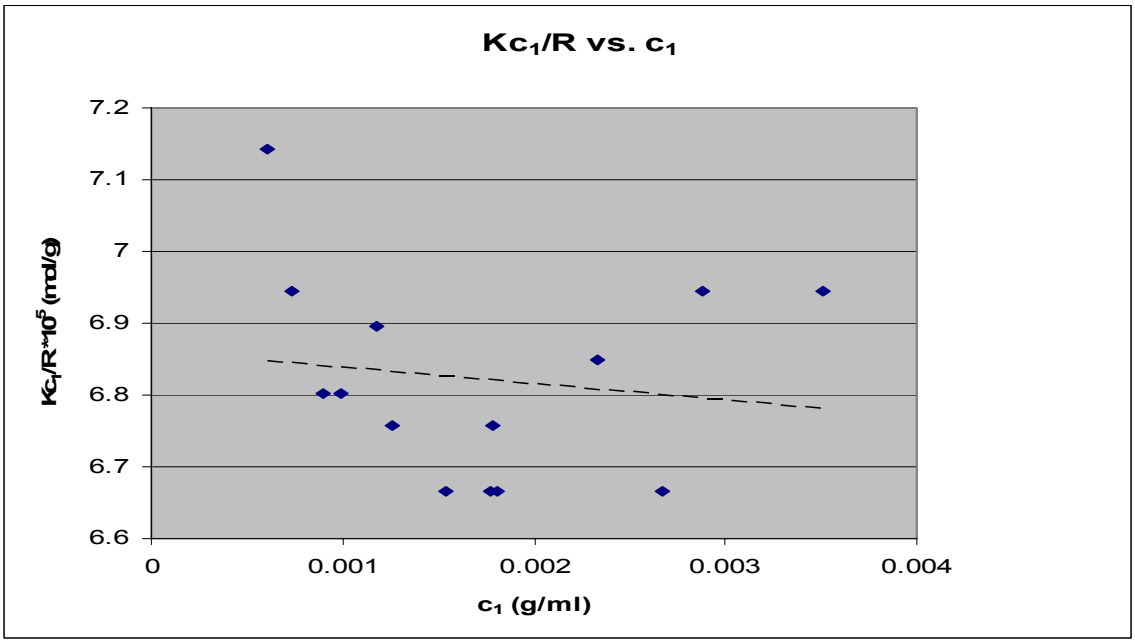

Figure 16. Zimm Plot of data from gradient # 2 of hysterises experiments

The scattering from the randomized gradient sequences has a high margin of error, moreover, the slope of the Zimm plot is negative, which contradicts positive values reported in the literature under similar solution conditions.

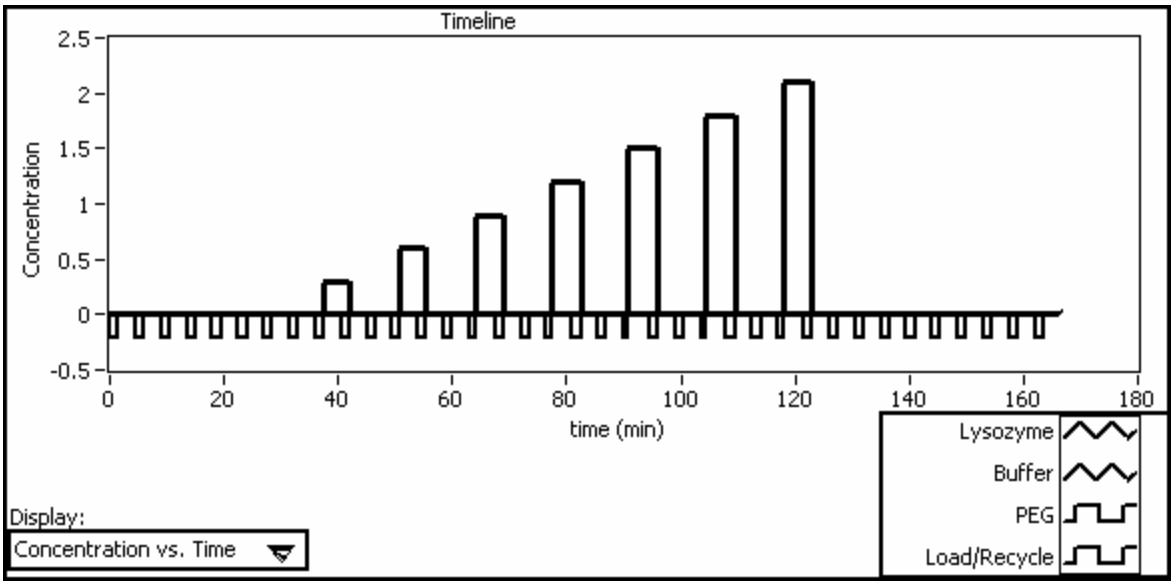

Figure 17. Illustration of gradient design # 3 of hysterises experiments

To check if the light scattering at a given concentration was being affected by previous concentrations in the gradient, the above gradient was designed by placing a 2 ml injection of pure buffer in between every concentration step of a seven step gradient.

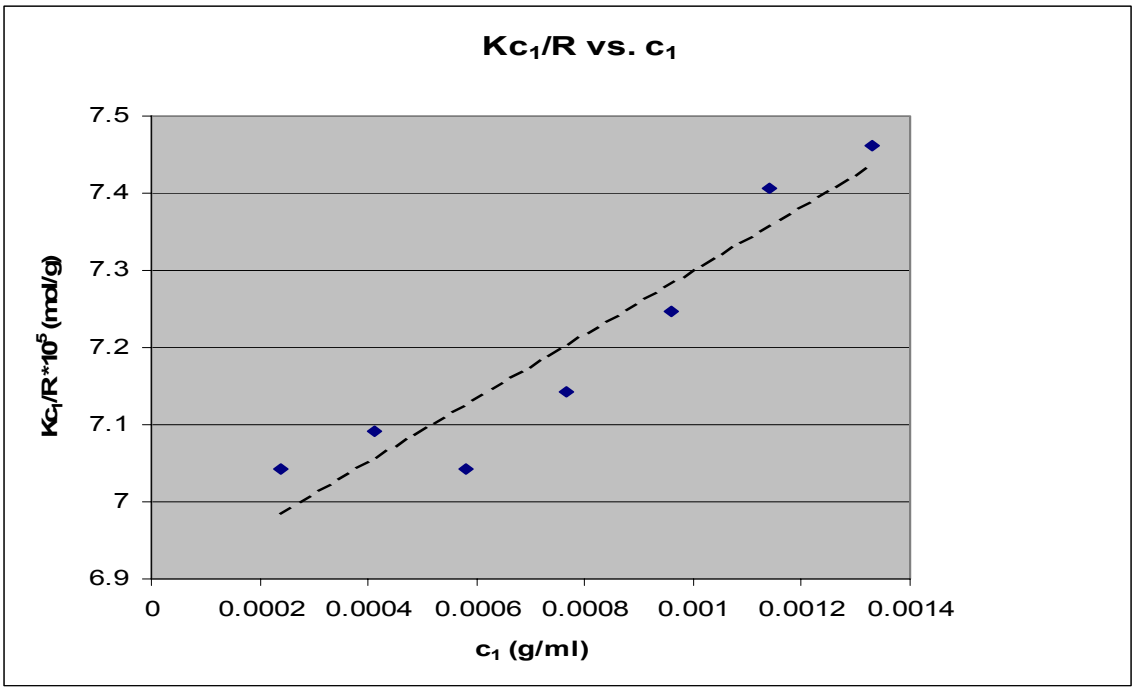

Figure 18. Zimm plot of data from gradient # 3 of hysterises experiments

The fit the scattering data was almost perfectly linear with the intercept yielding a molecular weight of 14285 g/mol, and a positive virial coefficient (slope). The plateau effect observed earlier above 1 mg/ml lysozyme is also completely absent. The data from this run suggests that a "blank" injection of pure buffer is necessary in between every concentration to minimize the margin of error in light scattering as well as to avoid the plateau in signals observed at higher concentrations.

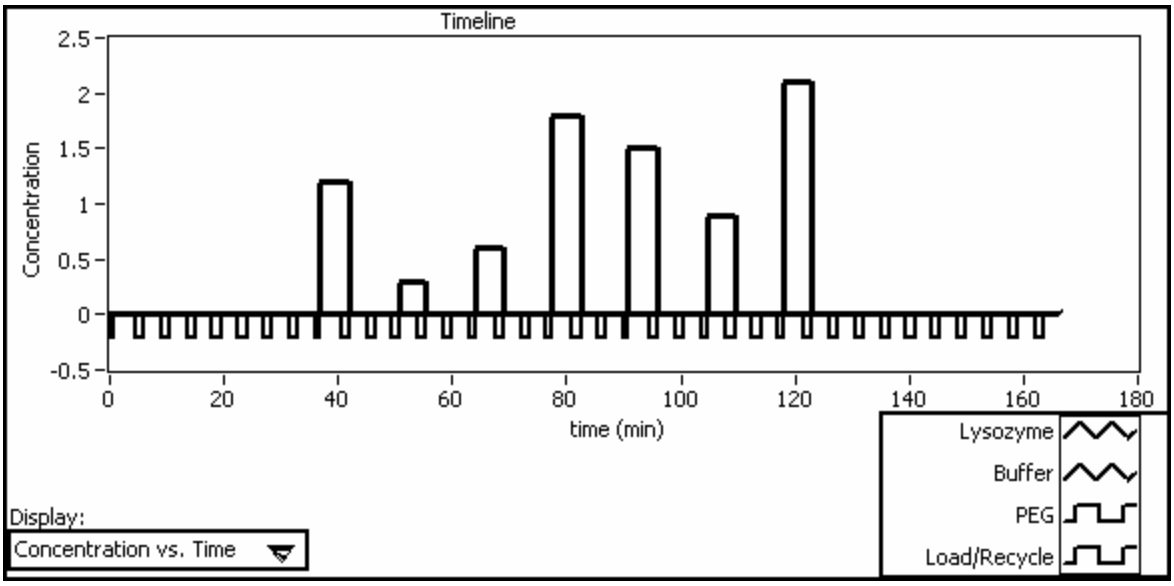

Figure 19. Illustration of gradient design #4 from hysterises experiment

The above gradient is exactly the same as the previous one except the order of the concentration steps is randomized. This gradient was designed to verify the results obtained from the previous run and to rule out any hysterises effects.

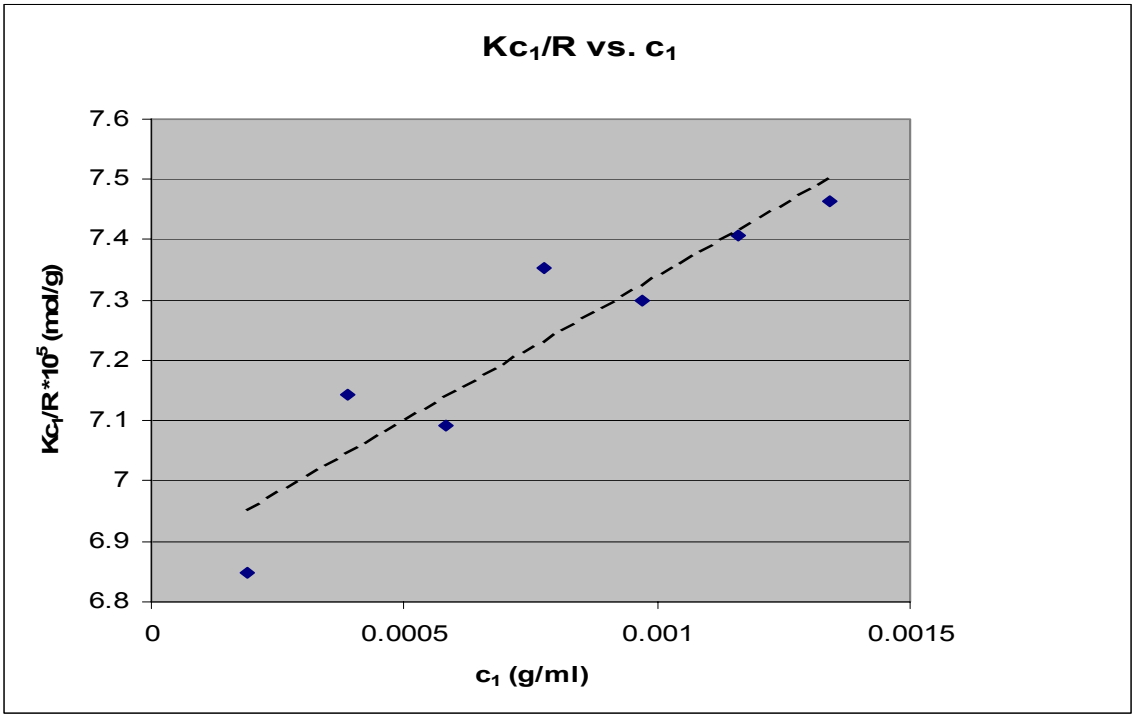

Figure 20. Zimm plot obtained from gradient #4 of hysterises experiments

The fit obtained to the data compares well against the fit from the previous run. The data confirms that a blank injection of pure buffer is necessary for a reliable fit.

4.4.2 Conclusions

The improvement in scattering obtained from the last two runs could be due to two factors: the lysozyme concentration range (0~2 mg/ml) being studied is too small for the detector sensitivity, resulting in heavy signal noise and variability in the data, and/or, the injection volume for each step (1 ml) in a continuous gradient is too small to flush out the previous concentration effectively, resulting in inconsistencies in scattering.

If the improvements obtained are entirely a result of the second reason and not due to instrument sensitivity, theoretically, increasing the injection volume at each step should

have the same effect as placing a 2 ml injection of pure buffer between concentrations, however, testing this hypothesis triples the amount of protein required per experiment and is not recommended.

4.5 Effect of PEG on Lysozyme Interactions

As stated in chapter 3, the second objective of this work is to characterize the effect of PEG 400 on lysozyme interactions in solution. Section 4.5.1 is a discussion of the theory adapted from Fraden et al.[2], while section 4.5.2 describes the experimental run designed around their theoretical model. Sections 4.5.3 is a summary of the results obtained followed by a discussion of the results in section 4.5.4.

4.5.1 Theory

For a two solute system, where 1 denotes protein and 2 denotes polymer, the excess Rayleigh ratio of the two solute system over that of the polymer solution alone is related to the protein concentration according to:

$$\frac{Kc_1}{(R_{1+2} - R_2)} = \alpha + \beta c_1 \tag{4.1}$$

α and β depend on the molecular weights, M_i, and differential refractive indices, n_i, of the added polymer, in addition to the polymer concentration.

$$\alpha = \frac{1}{M_1} + 4c_2 m B_{12} \tag{4.2}$$

$$\text{where, } m = \frac{M_2 n_2}{M_1 n_1} \tag{4.3}$$

$$\beta = 2\left[B_{11} + c_2 \left(3c_{112} - 2B_{12}{}^2 M_2\right) + mc_2 \left(3c_{112} + 2B_{11}B_{12}M_1\right)\right] \tag{4.4}$$

The first two bracketed terms in the above equation are collectively referred to the effective protein-protein virial coefficient, or B_{11}^{eff}. By measuring α and β as a function of c2, B_{11}, B_{12}, and C_{112} can be obtained. These values are then compared against their corresponding values calculated by assuming the molecules to be impenetrable hard spheres, obtained as outlined below:

$$B_{11}^{HS} = \frac{16\pi}{3} r_1^3 \tag{4.5}$$

$$B_{12}^{HS} = \frac{2\pi}{3}(r_1 + r_2)^3 \tag{4.6}$$

$$C_{112}^{HS} = \frac{8\pi^2}{27} r_1^3 \left(r_1^3 + 6r_2 r_1^2 + 15 r_1 r_2^2 + 8 r_2^3\right) \tag{4.7}$$

where, r_1 and r_2 are the effective hard sphere radii of the protein and polymer. For lysozyme, r_1 is approximated to be equal to its of 2.2 nm, whereas, r_2 is assumed to be equal to .6 nm as given by Kulkarni et al.[3]

4.5.2 Experimental Design

The experimental design used is based on the results obtained from the hysterises runs conducted earlier. Since it was found that an injection of 2 ml of pure buffer between concentrations of a composition gradient improves the quality of the fits obtained, the randomized design of hysterises circuit # 4 was used as a template for all the PEG runs. The only difference was a constant PEG concentration maintained throughout the run by syringe # 3. Thus, by measuring α and β as a function of c_1 for different values of c_2, B_{11}, B_{12}, and C_{112} were obtained. Measurements were made for a total of seven c_2 values, namely, 0 g/ml, .008 g/ml, .016 g/ml, .024 g/ml, .032 g/ml, .040 g/ml, and .048 g/ml.

The solution conditions were held constant throughout all seven runs at a pH of 4.65, and .02 M NaCl.

4.5.3 Materials and Methods

All glass-ware was thoroughly cleaned using warm water and phosphate-free detergent and dried for three hours, and then rinsed finally with Barnstead nanopure water. All solutions were made in sodium acetate buffer, pH 4.65 purchased from Sigma Aldrich. Lysozyme and PEG 400 were purchased from Fluka and Sigma Aldrich respectively. A 2mg/ml lysozyme stock solution was prepared by dialyzing the protein against large quantities of sodium acetate buffer with added NaCl to bring the total ionic strength to .02 M. The sample was then filtered through a .45 micron PVDF syringe filter and stored in clean glass vial and stored at 40 C. PEG 400 is a liquid at room temperature, with a density of 1.28 g/ml. A PEG stock solution at 80 mg/ml was made similar to the lysozyme solution and stored at 40 C. 200 ppm sodium azide was added to all stock solutions as well as buffer to prevent bacterial growth. All light scattering measurements were made at 90^0 on Wyatt Technology's miniDAWN TREOS. All protein concentration measurements were made using Varian Inc.'s ProStar 325.

4.5.4 Results

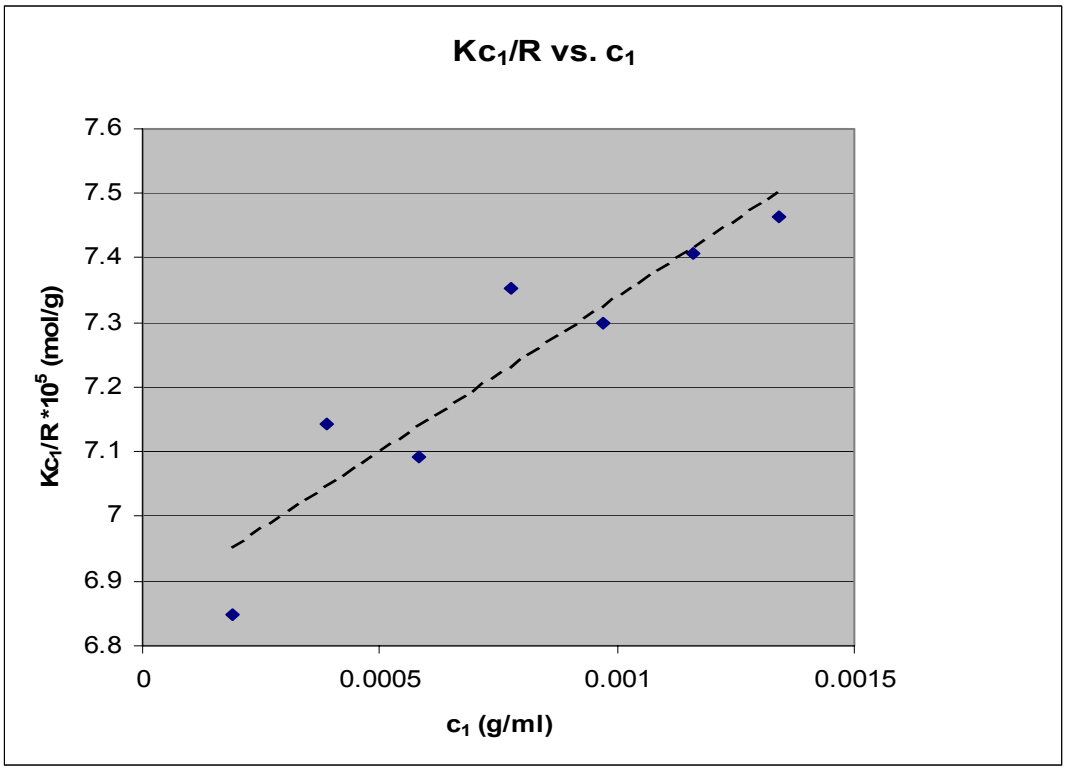

Fig. 21. Zimm plot from first run with c_2=0 mg/ml PEG; M_1 = 14285 g/mol; B_{11}= .0048/2 = 2.4*10^{-3} mol.ml/g^2

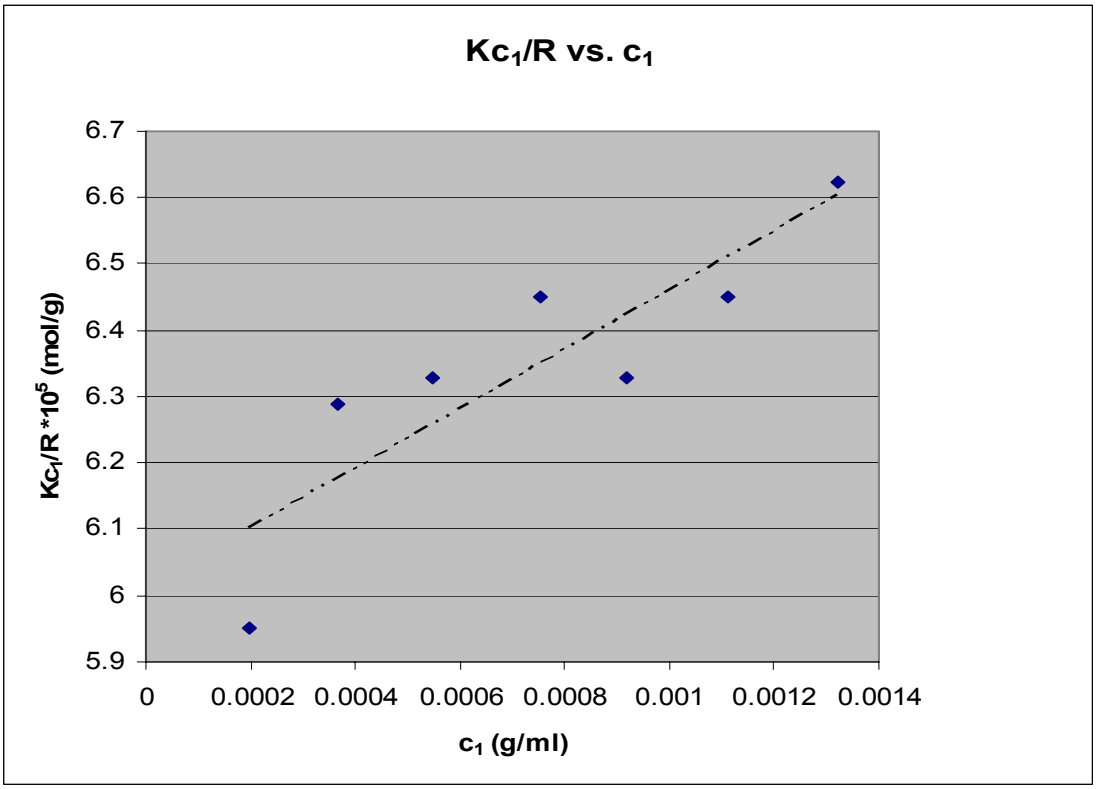

Figure 22. Zimm Plot from run #2 with c2=.008 g/ml PEG; $\alpha = 6*10^{-5}$ mol/g; $\beta = .0045$ mol.ml/g^2

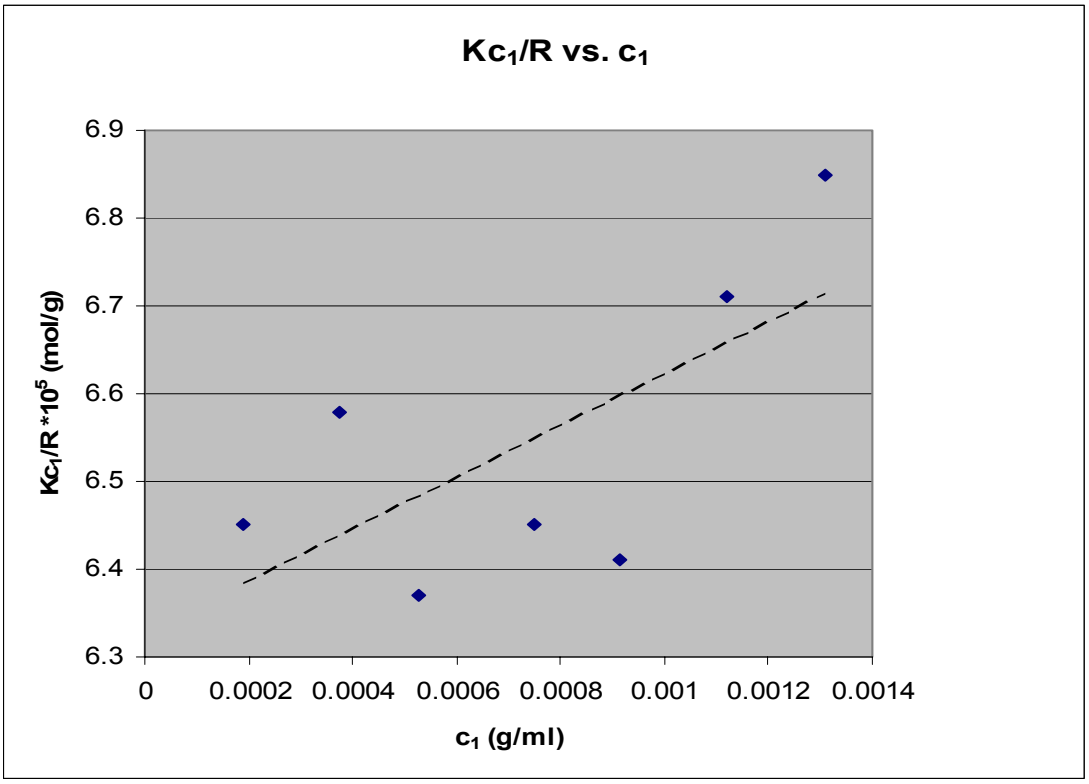

Figure 23. Zimm plot from run#3 with c2=.016 g/ml PEG; $\alpha = 6*10^{-5}$ mol/g; $\beta = .0029$ mol.ml/g^2

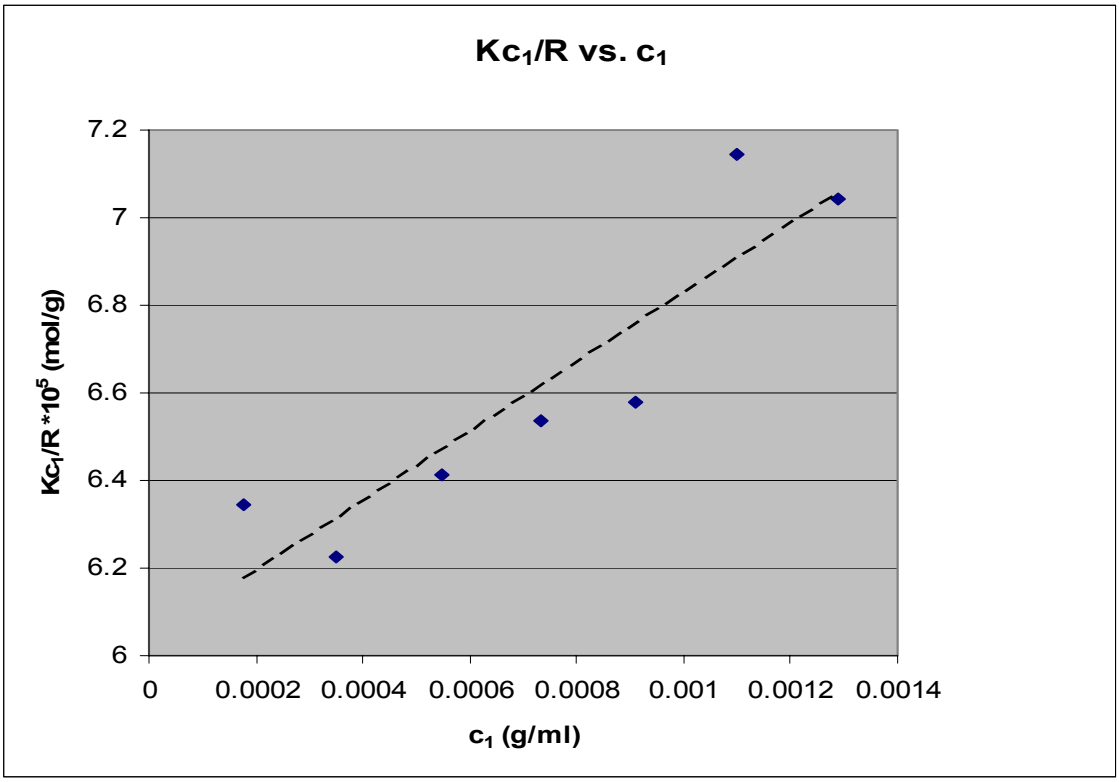

Figure 24. Zimm Plot from run #4 with c2 = .024 g/ml PEG; $\alpha = 5*10^{-5}$ mol/g; $\beta = .015$ mol.ml/g^2

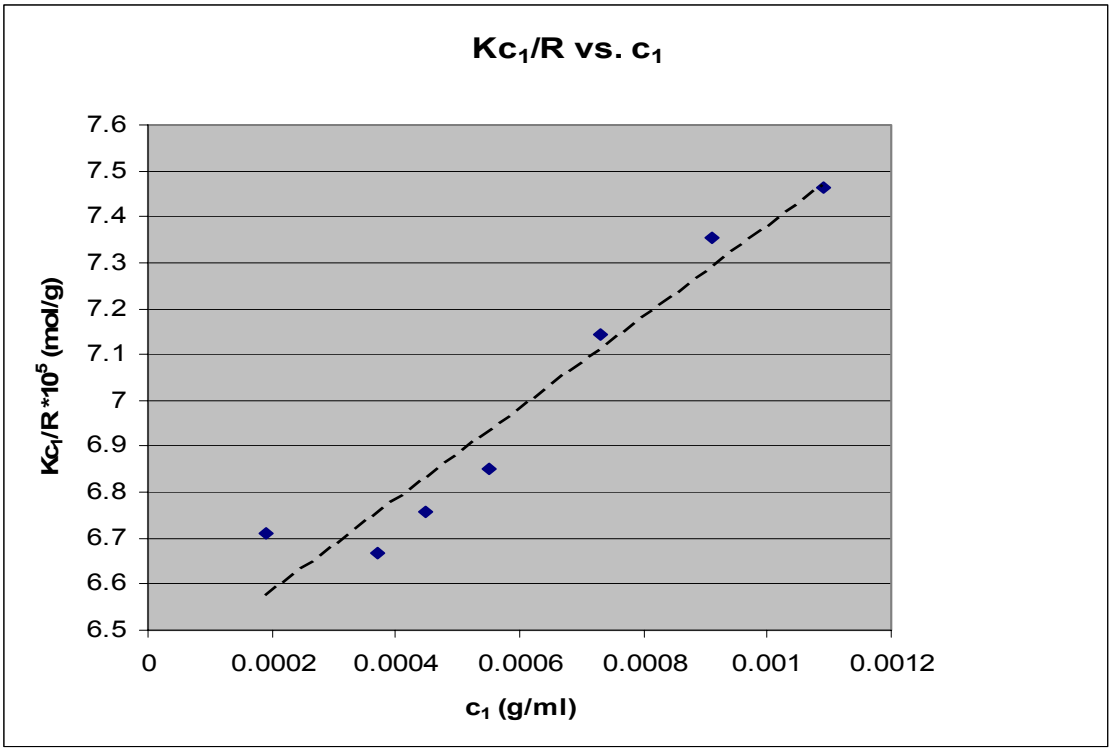

Figure 25. Zimm Plot from run #5 with c2 = .032 g/ml PEG; $\alpha = 6*10^{-5}$ mol/g; $\beta = .01$ mol.ml/g^2

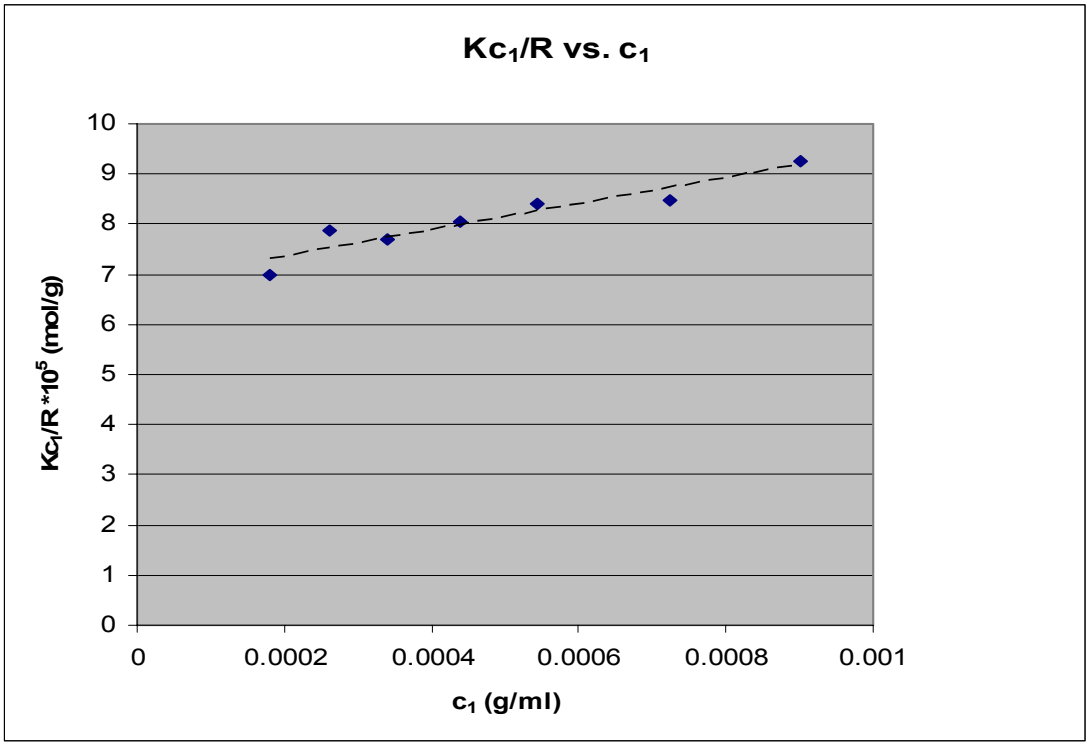

Figure 26. Zimm plot from run #6 with c2 = 0.040 g/ml PEG; $\alpha = 7*10^{-5}$ mol/g; $\beta = .026$ mol.ml/g^2

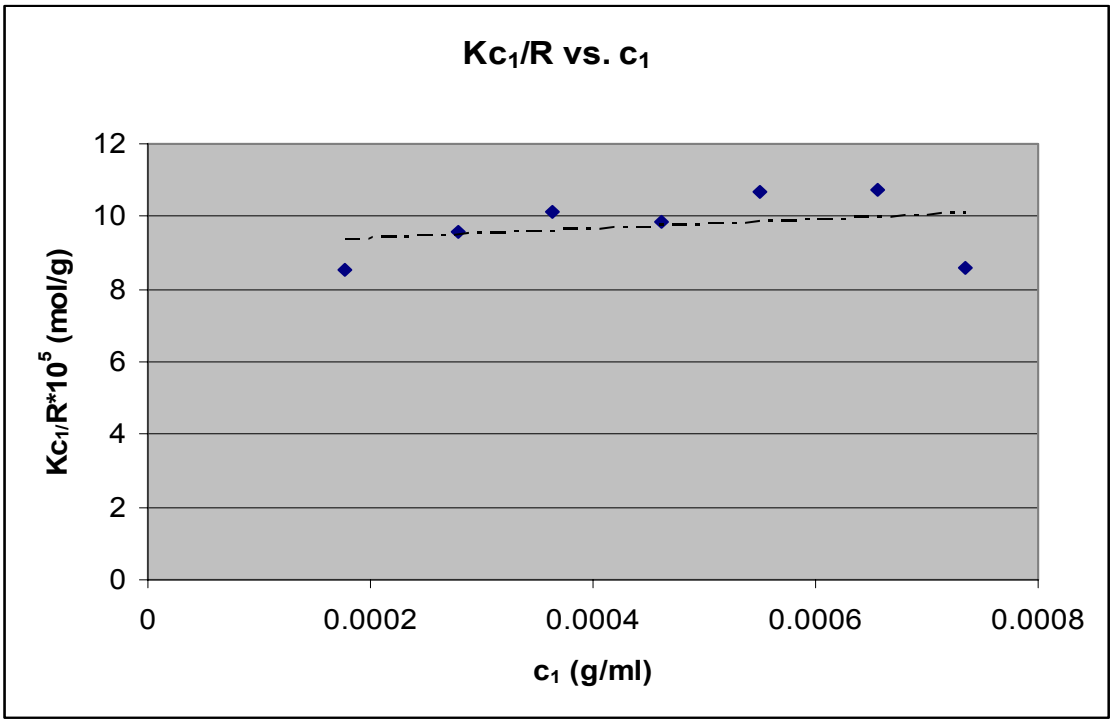

Figure 27. Zimm plot from run #7 with c2 = .048 g/ml PEG; α = .00009 mol/g; β = .0121 mol.ml/g^2

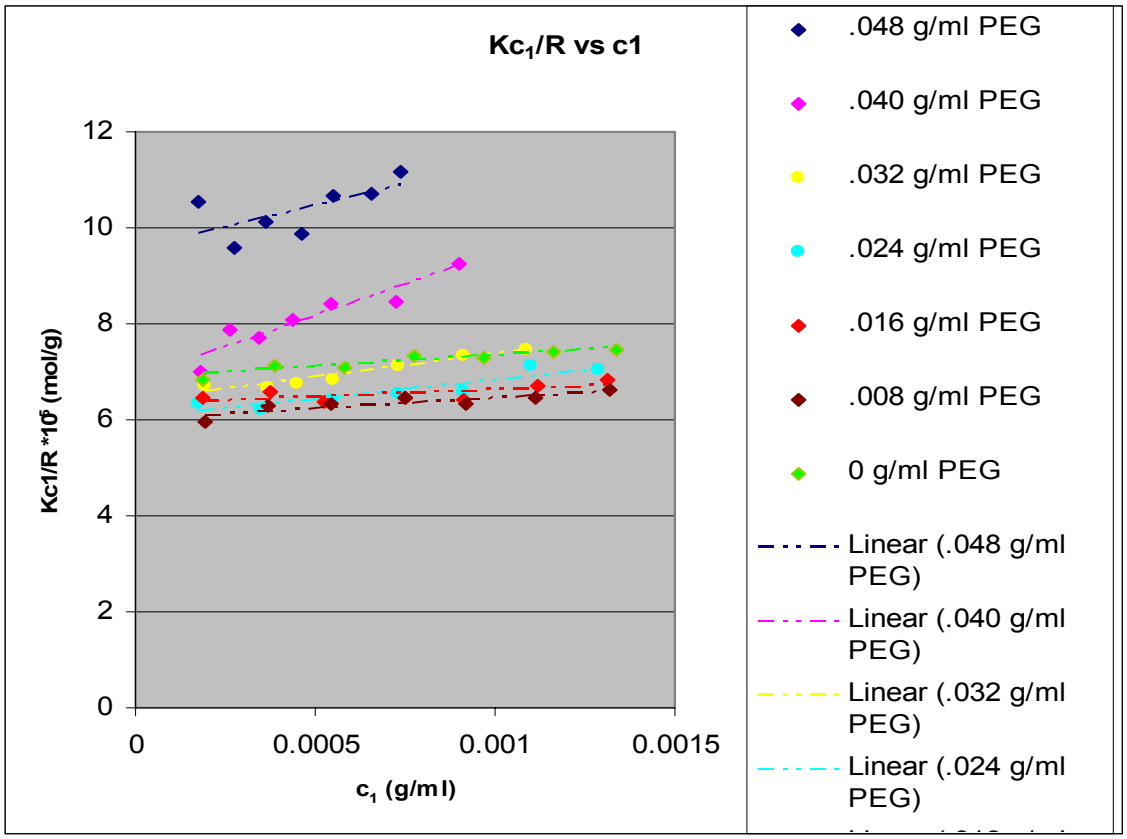

Figure 28. Comparison of Zimm plots obtained at seven different PEG concentrations

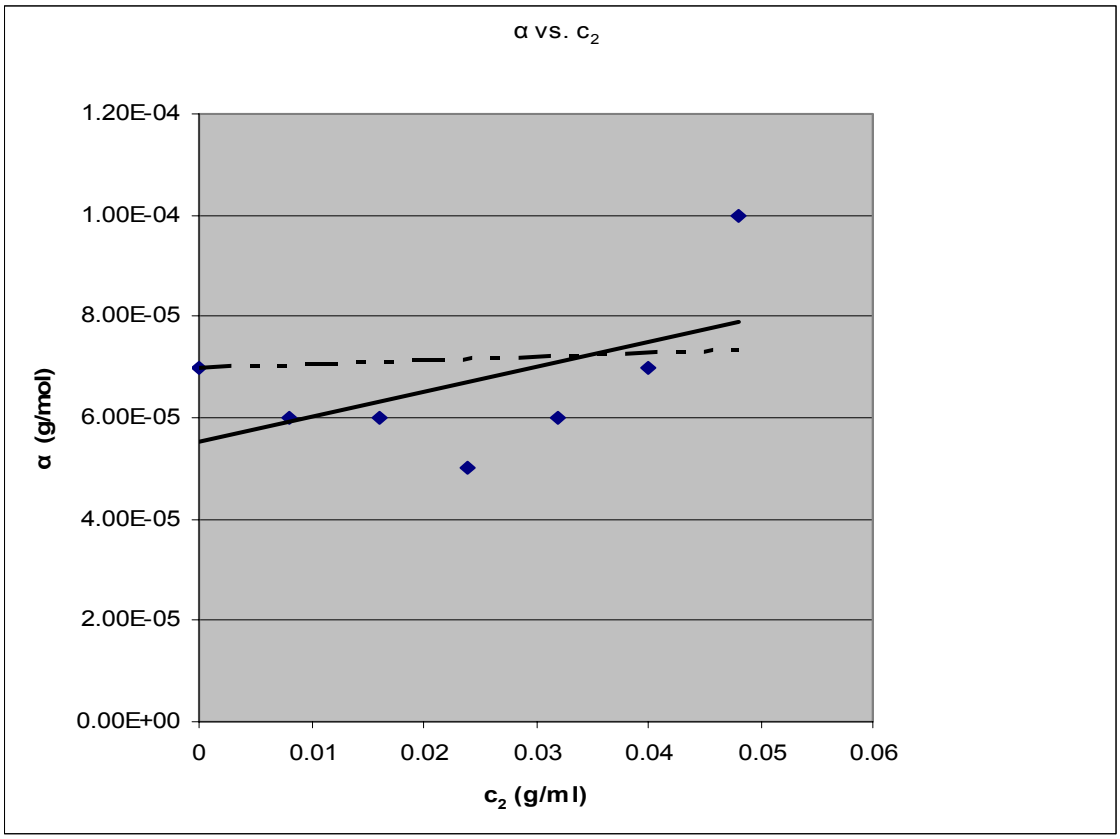

Figure 29. Plot of α as function of PEG concentration

The solid line is a linear fit to all the data points. However, as seen from the graph, the intercept obtained from this fit overestimates the lysozyme molecular weight compared to the result obtained from run#1. Nevertheless, $\partial\alpha / \partial c_2$ is still positive. The dashed fit is obtained by restricting the intercept to the molecular weight obtained from run#1, and the following are obtained:

From plotted data and equation 4.2,

$\frac{1}{M_1} = 7*10^{-5}$, and $4mB_{12} = 7*10^{-5}$;

$m = \frac{M_2 n_2}{M_1 n_1} = \frac{400 * 1.34}{14285 * 1.85} = .02028$; ($n_2$ and n_1 values were taken from Fraden et al.)

$\therefore B_{12} = 8.63 * 10^{-4}\, mol.ml / g^2$

Plotting β and c_2 according to equation 4.4:

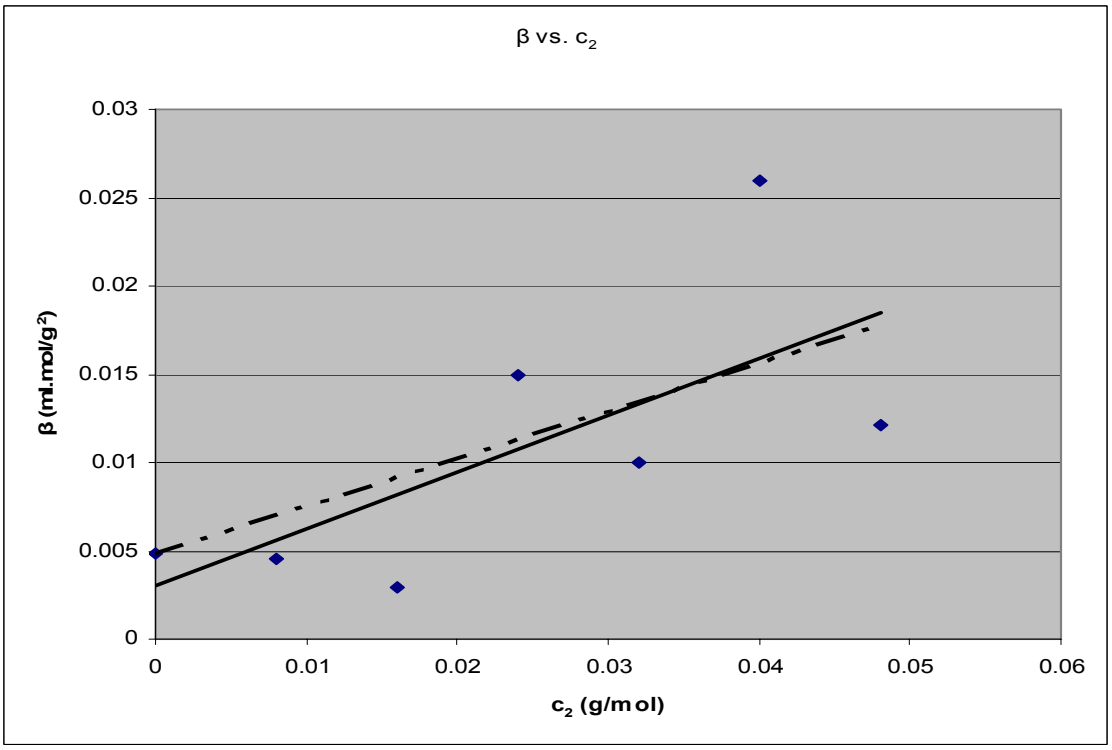

Figure 30. Plot of β as a function of PEG concentration

The solid line is a fit to the data according to equation 4.4, however, the B_{11} value obtained from the intercept of this fit is smaller than the B_{11} value obtained from run#1 with no PEG. The dashed fit was obtained by constraining the intercept to match the B_{11} value obtained in run#1. For comparison sake, values for C_{112}, and $\partial B_{11}^{eff} / \partial c_2$ were

obtained from both fits to check for any differences in the qualitative information conveyed by them.

From the intercept obtained through the solid fit and equation 4.4,

$B_{11} = 1.5*10^{-3}$ ml.mol/g^2;

From the solid fit and coefficients of c_2 in equation 4.4,

$$.3214 = 3(1+2m)C_{112} - 2B_{12}{}^2 M_2 + 4mB_{11}B_{12}M_1$$

Substituting values for B_{11}, B_{12} as calculated from α vs. c_2, m as calculated above, M_1 as obtained in run 1, and M_2 as given by the manufacturer,

$C_{112} = 103.37*10^{-3} ml^2.mol/g^3$, and, $\partial B_{11}^{eff} / \partial c_2 = 3C_{112} - 2B_{12}^2 M_2 = .3095$.

Repeating the above calculations using fit parameters from the dashed line,

$C_{112} = 86.13*10^{-3} ml^2.mol/g^3$, and, $\partial B_{11}^{eff} / \partial c_2 = .2577$.

Thus, the constraining of the fit to account for fluctuations in scattering does not alter the qualitative information conveyed by the data, with a positive C_{112} value and $\partial B_{11}^{eff} / \partial c_2 > 0$ in both cases.

The corresponding hard-sphere virial coefficients are calculated according to the following equations :

$$B_{12}^{HS} = \left(\frac{2\pi}{3}\right)(r_1 + r_2)^3 \text{; and, } C_{112}^{HS} = \left(\frac{8\pi^2}{27}\right) r_1^3 \left(r_1^3 + 6r_2 r_1^2 + 15 r_1 r_2{}^2 + 8 r_2^3\right).$$

Using r_2=2.2 nm (Fraden et al.), and r_2=.6 nm (Kulkarni et al.),

$B_{12}^{HS} = 4.84*10^{-3} ml.mol/g^2$, and, $C_{112}^{HS} = 5.76*10^{-3} ml^2 mol/g^3$.

4.5.5 Discussion and conclusions:

The protein-protein self virial coefficient obtained by Fraden et al is negative whereas, the coefficient obtained in this work is positive. However, that can be explained by the different solution conditions under which the two measurements were made. Measurements for the PEG 400 data set were made in sodium acetate buffer, pH=4.65, and .02 M NaCl, whereas, those for PEG 1000 were made in Sodium Phosphate buffer, pH= 6.2, and .2 M NaCl. At high ionic strengths such as .2 M NaCl, the virial coefficients tend to be negative due to complete screening of electrostatic repulsion between the protein molecules. The magnitude and sign of both the protein-polymer second and third virial coefficients are comparable for PEG 400 as well as PEG 1000. $\partial\alpha/\partial c_2$ is greater than zero for both molecular weights. As pointed out by Fraden et. al[2], the fact that $\partial\alpha/\partial c_2$ is a non-zero value indicates that scattering interference due to interactions between protein and polymer cannot be ignored, as assumed by Kulkarni et. al. Further, the calculated B_{12} values are smaller than the respective hard sphere values* for both molecular weights, indicating the presence of an attractive interaction between low molecular weight PEG and lysozyme that offsets the theoretical value based on hard sphere repulsion alone. A non-zero C_{112} value is obtained for both molecular weights, and is of the same order of magnitude as B_{12}, supporting Fraden et al.'s conclusion that a second virial term is not sufficient to account for the interactions between proteins in the presence of PEG, and a third virial term should be included in the model. Based on their model, Fraden et. al[2] summarize their observations in the form of a mathematical inequality as follows:

If $\partial B_{11}^{eff} / c_2 < 0$, PEG induces an attraction between protein molecules, whereas, $\partial B_{11}^{eff} / c_2 > 0$, PEG induces repulsion between the protein molecules. The inequality holds true for both PEG 400 as well as PEG 1000, indicating that low molecular weight PEGs influence protein interactions by inducing them to repel one another. The conclusion from light scattering data is confirmed by Fraden et al's[2] experiments studying the phase behavior of lysozyme in the presence of protein as a function of PEG concentration, where they showed that the cloud point for PEG 400 decresead with an increase in PEG concentration. A decrease in cloud point by the addition of PEG 400 indicates a "stabilization" of the liquid phase, contrary to the depletion effect that predicts an opposite behavior.

The lysozyme radius used by Fraden et al. to calculate the hard sphere virial coefficients is its radius of gyration, which they obtain by equating it to its hydrodynamic radius. However, since PEG is an uncharged molecule, the protein does not offer an electrostatic barrier to PEG and it can be expected that the protein allows an uncharged polymer a closer approach compared to another charged protein molecule. Moreover, Fraden et al.'s experiments are conducted at a fairly high salt concentration (.2 M), at which it can be expected that the r_1 should be smaller due to screening of the charges surrounding the protein, even if it is interacting with another lysozyme molecule. This raises a question of whether Fraden et al.'s conclusion that a different qualitative behavior than that predicted by pure depletion is observed for PEG influenced lysozyme interactions through light scattering indeed true ; or is the disparity between experimental results and the depletion model simply a result of over-estimating r_1+r_2?

To test the effect of a smaller protein radius on depletion, the hard sphere B_{12} and C_{112} values were calculated for both $r_1 = 2.2$ nm (Fraden et al.) and $r_1 = 1.8$ nm (Paliwal et al.). The results are summarized in table 6. It was found that considering a smaller protein radius did not make the depletion attraction disappear, i.e., $\partial B_{11}^{eff} / \partial c_2$ was less than zero for both cases, however, it did decrease by an order of magnitude for the smaller radius. The PEG radius was assumed to be equal to .6 nm (Kulkarni et al.) for both calculations. Thus, although the depletion attraction induced by PEG is a function of the hard-sphere radii considered, for all the values considered here, the interaction still remains attractive, contrary to the experimental results obtained by Fraden et al. for PEG 1000 as well the ones obtained in this work for PEG 400.

The following graph illustrates the reduction in depletion attraction by reducing the protein radius assuming that the polymer radius is constant.

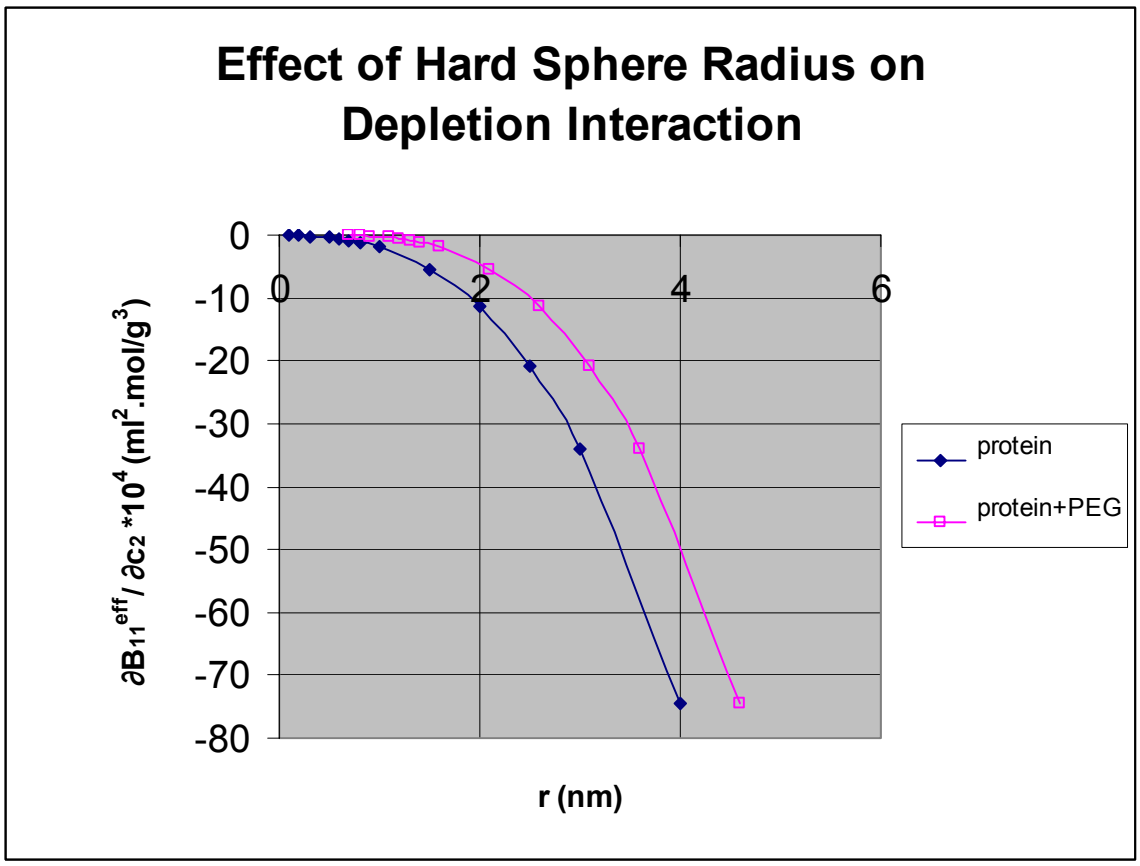

Figure 31. Effect of hard sphere radius on depletion interaction

As can be seen, the depletion interaction remains attractive throughout (y axis values are always negative), however, the magnitude of the attraction diminishes greatly at protein radii below ~ 1.5 nm.

In conclusion, the effect of low molecular weight PEG on protein interactions is characterized by the second and third mixed virial coefficients. The results obtained in this work as well as those obtained for PEG 1000 provide evidence that a depletion model may be insufficient to account for the interactions between protein molecules in the presence of low molecular weight PEG. Although the high degree of scattering

present in the data calls for more rigorous experimental control and data collection at PEG concentrations higher than 0.048 g/ml, as seen above, the qualitative information conveyed by all the fits remains the same.

Fraden et al.[2] attribute a possible attraction between protein and polymer to the PEG side chain forming hydrogen bonds with amino acid residues on the surface of the protein. Their theory, if true, could explain for the slightly higher molecular weight obtained from the solid fit in figure 29. Their proposition is one of the several propositions made in the literature[4,5] (see chapter 1) suggesting the presence of attractive interactions between protein and PEG, and this work provides experimental evidence supporting these claims for low molecular weight PEG. A more complete picture will emerge if this work is extended beyond PEG 400 to higher molecular weight PEGs as well, forming a ground for future research.

Furthermore, this work also introduces a novel flow technique for making light scattering measurements, conserving both the amount of protein as well as run time per virial coefficient measurement. Although it was found that the original composition gradient design introduced signal variability, a simple modification by inserting injections of pure buffer between each concentration significantly enhanced the data quality and reliability.

4.6 Referenes:

1. George, A., and W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallographica Section D.* 50:361–365.

2. . Bloustine, J., Virmani, T., Thurston, G., and Fraden, S. 2006. Light scattering and phase behavior of lysozyme-poly(ethylene glycol) mixtures. *Physical Review Letters.* 96:087803.

3. Kulkarni, A., Chatterjee, A., Schweizer, K., and Zukoski, C. 2000. Effects of polyethylene glycol on protein interactions. *Journal of Chemical Physics.* 113:9863-9873.

4. Sheth, S., and Leckband, D. 1997. Measurements of attractive forces between proteins and end-grafted (poly)ethylene glycol chains. *Proceedings of The National Academy of Sciences of The United States of America. 94:8399-8404.*

5. Abbott, N., Blankschtein., and Hatton, T. 1992. Protein partitioning in 2-phase aqueous polymer systems: a neutron scattering investigation of the polymer solution structure and protein-polymer interactions. *Macromolecules.* 25:3932-3941

APPENDIX

Run #	Detector flow-rate	Number of steps	Initial stock concentration
1	**low**	**low**	**low**
2	**high**	**low**	**low**
3	**low**	**low**	**high**
4	**low**	**high**	**low**
5	**high**	**high**	**low**
6	**high**	**high**	**high**
7	**low**	**high**	**high**
8	**high**	**low**	**high**

Table 1. Parameter levels in screening experiments

Tris buffer, .02 M NaCl, pH 7.5	
c1 (g/ml)	$Kc_1/R * 10^8$ (mol/g)
0.000187	7.81250
0.000439	8.06452
0.000702	8.40336
0.000956	8.62069
0.001190	8.84956
0.001430	8.92857
0.001680	9.09090
0.001920	9.17431
0.002170	9.17431

Table 2. Summary of data obtained using improvised parameters from screening experiments

Table 3. Data giving scattering as a function of lysozyme concentration for seven different runs, each with a different but constant value of PEG

c_2 = 0 g/ml PEG, 02 M NaCl, pH 4.65	
c_1 (g/ml)	$Kc_1/R * 10^5$ (mol/g)
0.000188	6.84931
0.000389	7.14285
0.000581	7.09219
0.000778	7.35294
0.000971	7.29927
0.001160	7.40740
0.001340	7.46268
c_2 = .008 g/ml PEG, .02 M NaCl, pH 4.65	
0.000196	5.95238
0.000366	6.28930
0.000546	6.32911
0.000753	6.45161
0.000919	6.32911
0.001110	6.45161
0.001320	6.62251
c_2 = .016 g/ml PEG, .02 M NaCl, pH 4.65	
0.000190	6.45161
0.000375	6.57894
0.000527	6.36942
0.000749	6.45161
0.000915	6.41025
0.001120	6.71140
0.001310	6.84931
c_2 = .024 g/ml PEG, .02 M NaCl, pH 4.65	
0.000177	6.34546
0.000352	6.22436
0.000549	6.41256
0.000732	6.53594
0.000911	6.57894
0.001100	7.14285
0.001290	7.04225

Table 3 continued on next page

Table 3 continued,

c_2 = .032 g/ml PEG, .02 M NaCl, pH 4.65	
0.000192	6.71140
0.000369	6.66666
0.000450	6.75675
0.000549	6.84931
0.000731	7.14285
0.000910	7.35294
0.001090	7.46268
c_2 = 0.040 g/ml PEG, .02 M NaCl, pH 4.65	
0.000181	6.99300
0.000261	7.87401
0.000341	7.69230
0.000443	8.06451
0.000544	8.40336
0.000725	8.47457
0.000921	9.25925
c_2 = .048 g/ml PEG, .02 M NaCl, pH 4.65	
0.000177	8.52546
0.000278	9.56547
0.000364	10.14521
0.000462	9.86021
0.000550	10.68376
0.000656	10.71811
0.000735	8.58761

c_2 (g/ml)	α (mol/g)	β (ml.mol/g2)
0	$7*10^{-5}$	.0048
.008	$6*10^{-5}$	.0045
.016	$6*10^{-5}$	.0029
.024	$5*10^{-5}$	.0150
.032	$6*10^{-5}$	.0100
.040	$7*10^{-5}$	.0260
.048	$9*10^{-5}$	.0121

Table 4. α and β as a function of c_2

M_1 (g/mol)	M_2 (g/mol)	B_{12} (mol.ml/g^2)	C_{112} (mol.ml^2/g^3)	B_{12}/B_{12}^{HS}	C_{112}/C_{112}^{HS}	$\partial B_{11}^{eff}/\partial c_2$
14285	400	$8.63*10^{-4}$	$86.13*10^{-3}$	.283	14.953	.2577

Table 5. Summary of results

B_{12}^{HS}	C_{112}^{HS}	$\partial B_{11}^{eff}/\partial c_2$ (hard sphere)
4.84*10-^{3}	5.76*10-^{3}	$-8.72*10^{-4}$

Table 6. Hard sphere virial coefficients

BIBLIOGRAPHY

1. Aldridge, S. 2006. Downstream processing needs a boost. *Bioprocessing Channels: Feature Articles.* 26:1.

2. Sommerfeld, S., and Strube, J. 2005. Challenges in biotechnology production – generic processes and process optimization for monoclonal antibodies. *Chemical Engineering and Processing.* 44:1123-1137.

3. Shenoy, B., Wang, Y., Shan, W., and Margolin, L. 2000. Stability of crystalline proteins. *Biotechnology and Bioengineering.* 73:358-369

4. Weber, P. 1991. Physical principles of protein crystallization. *Advances in Protein Chemistry*. 41:1-36.

5. George, A., and W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallographica Section D.* 50:361–365.

6. Cantor, C., and Schimmel, R. 1980. Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function. W. H. Freeman and Company, San Francisco.

7. McMillan, W., and Mayer J. 1945. The statistical thermodynamics of multicomponent systems. *Journal of Chemical Physics.* 13:276 –305.

8. Neal, B., Asthagiri, D., and Lenhoff, A. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophysical Journal.* 75:2469-2477.

9. Prasad, D., Virendra, U., and Rajiah, A. 1995. Dependence of the second virial coefficient on temperature. *The Chemical Engineering Journal and The Biochemical Engineering Journal.* 56:73-76

10. Vilker, V., Colton, C., and Smith, K. 1981. The osmotic pressure of concentrated protein solutions: effect of concentration and pH in saline solutions of bovine serum albumin. *Journal of Colloid Interface Science.* 79:548 –566.

11. Elcock, A., and McCammon, A. 2001. Calculation of weak protein-protein interactions: the pH dependence of the second virial coefficient. *Biophysical Journal.* 80: 613–625.

12. Dumetz, A., Snellinger-O'brien, A., Kaler, E., and Lenhoff, A. 2007. Patterns of protein-protein interactions in salt solutions and implications for protein crystallization. *Protein Science.* 16:1867-77.

13. Farnum, M., and Zukoski, C. 1999. Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophysical Journal.* 76:2716-2726.

14. Cantor, C., and Schimmel, R. 1980. Biophysical Chemistry: Part I: The conformation of biological macromolecules. W. H. Freeman and Company, San Francisco.

15. Rozhkov, S. 2006.Solubility and phase transitions in a water-protein-salt system. *Biofizika.* 51: 822-826.

16. Ruckenstien, E., and Shulgin, I. 2006. Effect of salts and organic additives on the solubility of proteins in aqueous solutions. *Advances in Colloid and Interface Science*. 123-126:97-103.

17. Cohn, E. 1925. The Physical Chemistry of the Proteins. *Physiological Reviews*. 5:349-437.

18. Roe, S. 2001. Protein Purification Techniques: A Practical Approach. Oxford University Press, New York.

19. Mcpherson, A. 2001. A comparison of salts for the crystallization of macromolecules. *Protein Science.* 10:418-422.

20. Kimber, M., Vallee, F., Houston, S., Necˇakov, A., Skarina, T., Evdokimova, E., Beasly, S., Christendat, D., Savchenco, A., Arrowsmith, C., Vedadi, M., Gerstein, M., and Edwards, A. 2003. Data-mining crystallization data-bases: knowledge-based approaches to optimize protein crystal screens. *Proteins: Structure, Function, and Genetics.* 51:562-568.

21. Lund, M., and Johnsson, B. 2003. A mesoscopic model for protein-protein interactions in solution. *Biophysical Journal.* 85:2940-2947.

22. Zhang, Y., and Cremer, P. Interactions between macromolecules and ions: the Hofmeister series. *Current Opinion in Chemical Biology.* 10:658-663.

23. Velev, O., Kaler, E., and Lenhoff, A. 1998. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophysical Journal.* 75:2682-2697.

24. Pjura, P., Lenhoff, A., Leonard, S., and Gittis, A. 2000. Protein crystallization by design: chymotrypsinogen without precipitants. *Journal of Molecular Biology.* 300:235-239.

25. Haas, C., Drenth, J., Wilson, W. 1999. Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution. *Journal of Physical Chemistry B.* 103:2808-2811.

26. Guo, B., Kao,S., McDonald, H., Asanov, A., Combs, L., and Wilson, W. Correlation of second virial coefficients and solubilities useful in protein crystal growth. *Journal of Crystal Growth.* 196:424-433.

27. Rupert, S., Sandler, S., and Lenhoff, A. 2001. Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnology Progress.* 17:182-187.

28. Liu, W., Bratko, D., Prausnitz, J., and Blanch, H. 2004. Effect of alcohols on aqueous lysozyme–lysozyme interactions from static light-scattering measurements. *Biophysical Chemistry.* 107:289-298.

29. Polson, A., Fotgeiter, G., Largier, J., Mears, G., and Joubert, F. 1964. The fractionation of protein mixtures by linear polymers of high molecular weight. Acta *Biochimica et Biophysica Acta.* 82:463-475.

30. Paliwal, A., Asthagiri, D., Abras, D., Lenhoff, A., and Paulaitis, M. 2005. Light-scattering studies of protein solutions:role of hydration in weak protein-protein interactions. *Biophysical Journal.* 89:1564-1573.

31. Arakawa, T., and Timasheff, S. 1985. Mechanism of poly(ethylene glycol) interactions with proteins. *Biochemistry.* 24:6756-6762.

32. Atha, D., and Ingham, K. 1981. Mechanism of precipitation of proteins by polyethylene glycol: analysis in terms of excluded volume. *The Journal of Biological Chemistry.* 256:12108-12117.

33. Ilett, S., Orrock, A., Poon, W., and Pussey, P. 1995. Phase behavior of a model colloid-polymer mixture. *Physical Review E.* 51: 1344-1352.
34. Tanaka, S., Ataka, M., Onuma, K., and Kubota, T. 2003. Rationalisation of membrane protein crystallization with polyethylene glycol using a simple depletion model. *Biophysical Journal.* 84: 3299–3306.

35. Kulkarni, A., Chatterjee, A., Schweizer, K., and Zukoski, C. 2000. Effects of polyethylene glycol on protein interactions. *Journal of Chemical Physics.* 113:9863-9873.

36. Baskir, J., Hatton, A., and Suter, U. 1986. Thermodynamics of the separation of biomaterials in two phase aqueous polymer systems: effect of the phase forming polymers. *Macromolecules.* 20:1300-1311.

37. Forcinit, D., and Hall, C. 1990. Theoretical treatment of aqueous 2-phase extraction by using virial expansions- a preliminary report. *ACS Symposium Series.* 419:53-70.

38. Abbott, N., Blankschtein, D., and Hatton, T. 1991. Protein partitioning in 2-phase aqueous polymer systems: novel physical pictures and a scaling thermodynamic function. *Macromolecules.* 24 : 4334-4348.

39. Abbott, N., Blankschtein., and Hatton, T. 1992. Protein partitioning in 2-phase aqueous polymer systems: a neutron scattering investigation of the polymer solution structure and protein-polymer interactions. *Macromolecules.* 25:3932-3941

40. Sheth, S., and Leckband, D. 1997. Measurements of attractive forces between proteins and end-grafted (poly)ethylene glycol chains. *Proceedings of The National Academy of Sciences of The United States of America. 94:8399-8404.*

41. Israelachvili, J. 1997. The different faces of (poly)ethylene glycol. *Proceedings of the National Academy of Sciences of The United States of America. 94:8378-8379.*

42. Bloustine, J., Virmani, T., Thurston, G., and Fraden, S. 2006. Light scattering and phase behavior of lysozyme-poly(ethylene glycol) mixtures. *Physical Review Letters.* 96:087803.

43. Lomakin, A., Asherie, N., and Benedek, G. 2003. Liquid-solid transition in nuclei of protein crystals. *Proceedings of The National Academy of Sciences of The United States of America.* 100:10254-10257.

44. Shaink, H., and Smit, J. 2000. Determination of the osmotic second virial coefficient and the dimerization of β-lactoglobulin in aqueous solutions with added salt at the iso-electric point. 2000. *Physical Chemistry, Chemical Physics.* 2:1537-1541.
45. Reˇsˇciˇc, J., Vlachy, V., Jamnik, A., and Glatter, O. 2001. Osmotic pressure, small angle x-ray, and dynamic light scattering studies of human serum albumin in aqueous solutions. *Journal of Colloid and Interface Science.* 239:49-57.

46. Haynes, C., Tamura, K., Korfer, H., Blanch, H., and Prausnitz, J. 1992. Thermodynamic properties of aqueous α-chymotrypsin solutions from membrane osmometry measurements. *Journal of Physical Chemistry.* 96:905-912.

47. Schafer, O. 1979. Modern aspects of membrane osmometer design. *Colloid and Polymer Science.* 257: 1007-1008.

48. Birdi, K. 2005. Membrane osmometry of aqueous micellar systems. *Colloid and Polymer Science.* 250:731-735.

49. Tessier, P., Lenhoff, A., and Sandler, S. 2002a. Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. *Biophysical Journal.* 82: 1620–1631.

50. Tessier, P., Sandler, S., and Lenhoff, A. 2004. Direct measurements of protein osmotic second virial coefficients by cross interaction chromatography.

51. Tangir, A., Ottens, M., Gijs, W., van Dedem, and Luuk., A. 2005. Design of self interaction chromatography as an analytical tool for predicting protein phase behavior. Journal of Chromatography A. 1089:111-124.

52. Lord Rayleigh. 1871. On the light from the sky, its polarization and color. *Philosophical Maganzine.* 41:107-120;274-279.

53. Debye, P. 1944. Light scattering by concentrated polymer solutions. *Journal of Applied Physics.* 15:338.

54. Zimm, B., 1948. The scattering of light and the radial distribution function of high polymer solutions. *Journal of Chemical Physics*. 16:1093.

55. Lyklema, J. 1991. Fundamentals of Interface and Colloid Science. Morgan Kauffman, San Francisco.

56. Wyatt., P. 1993. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta.* 272:1-40.

57. Kameyama, K., and Minton, A. 2006. Rapid quantitative characterization of protein interactions by composition gradient static light scattering. *Biophysical Journal.* 90:2164-2169.

58. Brown, W. 1996. Light scattering- principles and development. Oxford University Press, New York.